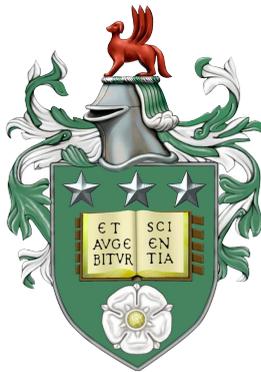# Travel behaviour modelling at the interface between econometrics and mathematical psychology

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy
by

Thomas Oliver Hancock

University of Leeds
July 2019

Intellectual property and publications

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 2 of this thesis has appeared in publication as follows:
Hancock, T.O., Hess, S., & 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.
I developed the main idea for this work, under the guidance of both Stephane Hess and Charisma Choudhury. I performed the modelling work and wrote the manuscript. Stephane Hess and Charisma Choudhury provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 3 of this thesis is a manuscript under review:
Hancock, T.O., Hess, S., & Choudhury, C. F. (under review). *A careful respondent or an uncertain response: Disentangling confounding sources of increased deliberation time using decision field theory.*
I developed the main idea for this work, performed the modelling work and wrote the manuscript. Stephane Hess and Charisma Choudhury provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 4 of this thesis is a manuscript under review:
Hancock, T.O., Hess, S., & Choudhury, C. F. (under review). *An accumulation of preference: contrasts between Decision Field Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for travel behaviour modelling.*
I developed the main idea for this work, performed the modelling work and wrote the manuscript. Stephane Hess and Charisma Choudhury provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 5 of this thesis is a manuscript under review:
Hancock, T.O., Hess, S., & Choudhury, C. F. (under review). *Quantum probability models: a new framework for modelling choices.*

I developed the main idea for this work, performed the modelling work and wrote the manuscript. Stephane Hess and Charisma Choudhury provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 6 of this thesis is a manuscript to be submitted shortly:
Hancock, T.O., Hess, S., & Choudhury, C. F. (in preparation). *New models for dynamic choice contexts: further steps towards bridging choice modelling with mathematical psychology.*
I developed the main idea for this work together with Stephane Hess. I performed the modelling work and wrote the manuscript. Stephane Hess and Charisma Choudhury provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 7 of this thesis is a manuscript to be submitted shortly:
Hancock, T.O., Hess, S., Daly, A., & Fox, J. (in preparation). *Improving forecasts and behavioural insights by applying model averaging across multiple choice models.*
I developed the main idea for this work together with Stephane Hess and Andrew Daly. I ran the models and wrote the manuscript together with Stephane Hess. All authors provided recommendations on the modelling. The manuscript was improved by comments from all the co-authors.

Acknowledgments

First and foremost I would like to say a big thank you to my supervisors, Stephane Hess and Charisma Choudhury, who have both been incredibly supportive from my first day at Leeds. Their guidance, knowledge and unfaltering belief in me has undoubtedly been the key component to me producing this thesis- I simply could not have wished for a better pair of supervisors.

I would also like to extend my thanks to my co-authors Andrew Daly and James Fox, who provided me with very beneficial alternative perspectives which helped me think outside the box.

Additionally I would like to thank the great many individuals who have given me feedback through reading my work or discussing it with me after conference seminars. I will not be able to name you all but in this regard I would particularly like to thank Caspar Chorus, Jerome Busemeyer and most especially Tony Marley, all of whom gave very insightful as well as positive feedback.

I am similarly grateful for the informative discussions I have had with Thijs Dekker, Matthew Beck, Danny Campbell and Seda Erdem over the course of completing this thesis. In particular, I would like to thank my examiners, Gerard de Jong and Michel Bierlaire, for a very interesting discussion over the course of my Viva, which as well as helping me identify the strengths and weaknesses of my work, helped direct me towards many avenues for future work. Thanks to Gerard for his feedback after my transfer, as well as encouraging me to be more ambitious!

My work has also benefited from both the feedback after and the experience itself of giving seminars, so a big thank you to Martina Vandebroek, Faisal Mushtaq and Mark Mon-Williams, and also Peter Hancock for inviting me to give seminars (without doubt the most surreal moment of my PhD was giving a seminar on quantum theory to my own father and his research group!)

I have additionally been lucky enough to have had several really interesting datasets given to me over the course of my PhD, so I would like to say a

big thank you to Andrew Cohen, Romain Crastes dit Sourd, Maria Börjesson and Caspar Chorus.

I have also greatly benefited from discussions with fellow CMC members (thanks to you all). In particular, I would like to thank Chiara Calastri, Romain Crastes dit Sourd, Martyna Bogacz and David Palma for not only your feedback but your constant support and friendship.

I would like to thank my brother Robert, cousin Michael as well as my friends Tom, Geraint and James for their support and encouragement. Thanks also to friends at both Acomb and Leeds University Quaker meetings for their moral support.

I owe a big thank you to my mum, dad and stepmum for all that they have done for me and their unfaltering support and belief in me. Thanks also to my Aunt Sarah, who convinced me to embark on a PhD, as well as my parents in law for their generous hospitality during my many writing retreats at their house over the last year or so!

Finally, a huge thank you to my wonderful wife Holly, for her constant love and support.

Abstract

Whilst the field of choice modelling has been dominated by approaches based on random utility maximisation (RUM), there has recently been a considerable rise in work considering alternative model structures that attempt to incorporate behavioural insights from psychology, behavioural science and other disciplines. Thus far, most alternative models have only involved small steps from the firm foundations of RUM models. However, a key issue with models not based on RUM is the loss of its micro-economic foundations and hence the ability to conduct welfare analysis. This thesis puts forward the key argument that if we are to lose this benefit, a newly proposed model needs to allow for rich account of behaviour and steps should be taken to move further from the tried and tested. Given that choice models developed within mathematical psychology have been specifically designed to explain contextual effects where alternatives impact each other, it is surprising that there has not yet been more of a bridge between the disciplines. This is hence the key aim of this thesis: to build bridges between the disciplines by bringing ideas and models from mathematical psychology into choice modelling in the context of travel behaviour research. We provide a large number of underlying methodological improvements and adaptations for two accumulator models, decision field theory (DFT) and the multi-attribute linear ballistic accumulator (MLBA). The work in this thesis provides thorough and detailed applications of both models to a wide range of travel behaviour choice contexts, such that the precise nature of the models can be established and the models can be contrasted to standard choice modelling methods. Furthermore, we establish best practices for the implementations of both models. Crucially, tests across a wide variety of choice scenarios demonstrate that these models regularly outperform standard choice models in terms of model fit, as well as providing useful behavioural insights. We also develop new frameworks for choice models implementing quantum logic, which has made a successful transition into cognitive psychology but has not yet been discussed in depth in the context of travel behaviour research. Results from applications of quantum models suggest that they provide an accurate account of changes in choice context. Of course, adding yet more possible models to the mix creates further issues for analysts in deciding which struc-

ture to adopt. With this in mind, we discuss in depth the concept of model averaging, demonstrating its potential within the context of travel behaviour research and showing how it can be applied across a wide variety of models to generate interesting insights by combining evidence across models. Overall, we demonstrate that if we are to step away from the firm foundations of RUM, the move towards models developed in mathematical psychology will provide considerable benefits for those who make the leap.

Contents

Contents

Contents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1 Background

Given that a fundamental component of human nature is the ability to deliberate on our decisions and make choices, it is unsurprising that researchers from psychology, philosophy, econometrics, behavioural science and neuroscience, to name a few, are all interested in the study of choice behaviour. Given the wide variety and nature of these different disciplines, it is understandable that the methods and types of applications used to study choice behaviour are hugely varied. In particular, the prediction or statistical analysis of choices has also been studied from very different viewpoints which is indicative of the abstract nature of attempting to put numbers on choices. As a result, the mathematical methods used by psychologists are often structurally completely different to traditonal choice modelling methods. What is more surprising is the lack of a bridge between these two different schools of thought given that they often attempt to model choices that are made under similar conditions.

In econometrics, the dominant theory for over forty years has been random utility maximisation (RUM), which started with the derivation of the logit formula (Luce, 1960) and work by Marschak (1960), before the well-known contribution of McFadden (1974). A key feature of this family of models is the notion of trade-offs, where good performance on one attribute can compensate for poor performance on another attribute. The utility or attractiveness of alternative is a function of the attributes of that alternative alone and is obtained through a single calculation involving all relevant attributes.

As a contrast, typical approaches in mathematical psychology have focussed on what are known as 'sequential sampling models' or 'accumulator models' (Busemeyer and Townsend, 1992; Ratcliff, 1978; Usher and McClelland, 2001), where the preference (their equivalent of a utility) for the different alternatives updates over time within a single choice context. A decision-maker samples the world, deliberates on the relative merits of the

different alternatives before choosing one. Consequently these models have also been called 'process models' as they attempt to mathematically represent the decision-making process as well as attempting to capture the choice outcomes. The result of this approach has lead to the development of models that can predict response times as well as choice behaviour (Brown and Heathcote, 2008). As such, one possible conclusion would be that, relative to econometricians, mathematical psychologists are focussed on *choice processes*, as whilst they may often study the results of responses to similar multi-alternative, multi-attribute choice tasks, it is rarely the preferences of an individual or their precise response to particular attributes of alternatives that psychologists are interested in. Instead, they focus more on how and why we make choices rather than what the choices actually are. Consequently the explanation of various contextual effects is often of key importance.

This crucial difference in effect is the result of a very different response to the findings from behavioural economics, with mathematical psychologists often focussing on ensuring that their models can account for and predict such behaviour. As a contrast, whilst mainstream choice modellers have attempted to add behavioural elements to their models, the computational ease of calculating and measuring trade-offs remains of key importance, meaning that there has been a reluctance to step away from traditional model structures. There has, however, been some attempts to incorporate new ideas. For example, the incorporation of regret has seen a recent rise of models based on a random regret minimisation (RRM) framework (Chorus, 2010; Chorus et al., 2008). However, a key theme of this thesis is the argument that if we are to incorporate new behavioural theories into our models at the cost of the loss of the ability to perform welfare analysis, then the model needs to be more fundamentally different and provide a more detailed account of behaviour. Indeed, RRM for example is still a logit model, albeit a non-RUM consistent one, and shares the fact that the regret of an alternative is calculated in a single process, rather than through some accumulation of evidence.

Models developed in mathematical psychology thus provide a key avenue for exploration, given the wide range of behaviours they have been used to explain. Additionally, as they have originated from a very different school of thought, they have very different underlying structures meaning they are considerably larger departures from the tried and tested. However, thus far, models from mathematical psychology have not made the transition into mainstream choice modelling. A key reason for this due to the fact that these models make very different assumptions about the effects that the different alternatives have on each other. Whereas the utilities of alternatives under random utility maximisation are specifically not impacted by the attributes of other alternatives, the precise converse is typically true in models developed in mathematical psychology, which is a naturally unappealing property

if one is to measure trade-offs. However, if we accept the loss of this feature, as we also do in RRM, allowing alternatives to have an impact on each other allows for a great deal of flexibility leading to a rich account of contextual behavioural effects. For example, similarity, compromise and attraction effects can be explained by decision field theory (Berkowitsch et al., 2014; Hotaling et al., 2010; Roe et al., 2001), the multi-attribute linear ballistic accumulator (MLBA) model (Trueblood et al., 2014) and other such models. These two particular models are perhaps more tractable than some of their counterparts from mathematical psychology, such as the leaky competing accumulator (Usher and McClelland, 2001), which relies entirely on simulation to calculate the probabilities for choosing alternatives. A large part of this thesis thus considers the operationalisation of these two models into mainstream choice modelling.

Decision field theory (DFT), first developed by Busemeyer and Townsend (1992), is a stochastic dynamic model of preference accumulation, making it very different from the standard random utility maximisation model. It assumes that a decision-maker randomly attends to a specific attribute at each moment and updates the preference values accordingly by comparing the alternatives across this attribute. The decision-maker then considers the different attributes for some number of preference updating steps until the preference value for an alternative (or preference difference between alternatives) either reaches some internal threshold, or the decision-maker reaches some external threshold (such as running out of time in which to further deliberate on the alternatives). MLBA is also very different (to both RUM and DFT), as it is not stochastic but is similarly dynamic. Under MLBA, each alternative has some random initial preference before drifting linearly (based on some evaluation of a 'drift rate' for the alternative) towards an internal threshold. The first alternative to reach this threshold is then chosen. Crucially, many different forms of both DFT and MLBA are possible, resulting in the fact that best practices for the implementation of such models are not known.

A further key benefit in the potential move to accumulator models is that they provide a more natural and intuitive framework in the context of dynamic choice settings, where the attributes of the alternatives change over time. Whilst dynamic discrete choice models (see e.g Cirillo and Xu 2011) have made a significant impact in the study of such behaviour, they evaluate the underlying process by considering some number of discrete intervals and estimate the utilities of alternatives at each interval, relying on of state dependence or lagged variables to create links between the moments. As a contrast, an accumulator model can estimate probabilities at different moments with a single continuous process that does not require individual moments to be 'stitched together'. Computational difficulties have thus far limited the application of accumulation models in the context of changing

attributes. One notable exception is Holmes et al. (2016)'s application of the
linear ballistic accumulator (the model that preceded MLBA, which simply
estimates a constant for the drift rate for each alternative) in the context of
dot motion perception tasks where the direction of the dots changed whilst
the decision-maker was deliberating on the task. They demonstrated that
the model could be adapted into a 'piecewise' accumulator such that chang-
ing information could be taken into account. However, the type of dynamic
behaviour that is typically examined in the travel behaviour community is
far more complex, with for example, the decision as to when to merge lanes
whilst driving subject to a large number of factors (Choudhury et al., 2009;
Kondyli and Elefteriadou, 2012).

Another avenue for exploration of bridging the gap between mathematical
psychology and mainstream choice modelling lies in the recent rise of models
based on quantum logic in cognitive psychology. Quantum mechanics, first
developed for use in experimental physics, may sound like a rather exotic
concept to consider for travel behaviour analysis but no more so than gravity
models. The key difference between classical and quantum probability is that
the distributivity law of probability $(A(B + C) = AB + AC)$ fails to hold
in quantum theory. This allows for a flexible approach in the explanation
of ordering effects (Trueblood and Busemeyer, 2010; Wang et al., 2014) and
interference effects (Aerts, 2012). Ordering effects, to which choice modellers
are no strangers, can typically be controlled for through careful experimen-
tal design, but the impacts of these effects remains a key interest to choice
modellers (Bateman et al., 2008; Carlsson et al., 2012; Nguyen et al., 2015).
It is therefore unsurprising that models based on quantum logic have begun
to appear, with the first papers on quantum models appearing in the Journal
of Choice Modelling (Lipovetsky, 2018) and Transportation Research Part
B (Yu and Jayakrishnan, 2018). It however remains to be seen whether the
full applicability of such approaches means that they will rival traditional
methods.

Whilst there has been some comparisons of models developed in mathe-
matical psychology with traditional methods such as the work of Berkowitsch
et al. (2014), rigorous tests of the models in the context of travel behaviour
modelling have not yet been conducted. Furthermore, a through understand-
ing of the the precise nature of the models developed in mathematical psy-
chology will help improve the transferability of these models into the sort of
choice modelling applications that econometricians typically perform. This
thus is the key aim of this thesis: to build bridges between mathematical psy-
chology and choice modelling through the careful examination and adaptation
of models developed in mathematical psychology for use in travel behaviour
choice contexts.

2 Research Gaps

As discussed in the previous section, whilst there has been a vast amount of work in parallel streams of choice behaviour research, a crucial key theme in the literature is in the lack of effort to bridge the gap between mainstream choice modelling and mathematical psychology. The work in this thesis attempts to bridge this divide by considering a number of steps for bringing models and ideas from mathematical psychology into mainstream choice modelling applications. We attempt this by considering the following series of key research gaps.

Gap 1: Tractable methods for the applications of accumulator models.

The first key gap is that there is a lack of clear, tractable methods for the application of accumulator models in mainstream choice modelling scenarios. Whilst random utility models can be applied to datasets with thousands of alternatives, the computational complexity of accumulator models has thus far prohibited their use (Otter et al., 2008; Trueblood et al., 2014; Tsetsos et al., 2010). A key issue lies in the estimation of a likelihood function, with many models developed in mathematical psychology relying on simulation to calculate probabilities even in basic choice scenarios (Krajbich et al., 2012; Usher and McClelland, 2001). A further issue is that models for which a likelihood function does exist, such as DFT and MLBA, tend to have a number of distinct 'process' parameters, for which the impacts in the context of basic choice data are unclear. Furthermore, for most models in mathematical psychology, a number of different versions of the models exist, with it often not being clear what the 'standard' approach should be. This complicates the development and expansion of such models, as there is often not a clear basis from which to start.

Gap 2: Rigorous comparisons of alternative approaches

The second gap that we consider is the lack of rigorous comparisons of models developed in mathematical psychology compared to standard choice models. A notable exception here is provided by (Berkowitsch et al., 2014), who tested DFT against logit and probit models. However, this application considered only basic implementations of all models, not testing for heterogeneity across decision-makers nor the relative abilities of the models to capture core concepts such as underlying preferences towards an alternative. Additionally, models from mathematical psychology have rarely been applied in the context of travel behaviour research, or tested on revealed preference datasets.

Gap 3: Developing a choice modelling framework based on quantum logic

The third gap considered is the lack of a clear framework for the adoption of quantum logic into choice models. This is mainly a result of the fact that quantum models within cognitive psychology are simply used to examine single or pairs of decisions made by participants under experimental conditions. Whilst Lipovetsky (2018) develops quantum models for marketing choices, it remains unclear how to apply quantum models to multi-attribute, multi-alternative choice, with Yu and Jayakrishnan (2018) emphasising that there could be numerous approaches for this issue. Additionally, it is unclear as to what context such approaches would be most suitable.

Gap 4: Identification of contexts for which accumulator and quantum models are suitable

Whilst the adoption of the various models discussed above may lead to richer behavioural insights, there is still the key drawback that these models will lose the ability to perform welfare analysis, which may prohibit use of these models even if the gains in model fit are more substantial than previous departures from traditional methods. Thus a key research gap for the development of such approaches is to understand the contexts in which these models will be most suitable. For example, Chorus (2015) notes that there has of yet been little discussion on the appropriateness of different models in moral choice contexts. With quantum models performing well in such scenarios in cognitive psychology, further research could test whether models with a quantum framework are also particularly suitable for moral contexts in larger scale choice modelling applications. Furthermore, accumulation models, as discussed in Section 1, have almost exclusively been tested in static choice scenarios only. Given that they provide a natural framework for the inclusion of changing attribute values, this may be where choice modellers stand to gain most from the adoption of these models, i.e. when the focus is also more on prediction than valuation.

Gap 5: Testing model averaging and combining approaches from different disciplines

As well as a lack of work contrasting and comparing accumulator models with standard choice models, there has thus far been very little work considering the benefits of approaches utilising both frameworks simultaneously. For example, the increasing interest in the travel behaviour community for the adoption and consideration of different decision rules has led to the implementation of latent class models with differing decision rules in the different

classes (Hess et al., 2012). The scope for increasing the variation in the decision rules improves considerably with the adoption of models with very different underlying structures such as accumulator or quantum choice models. The simultaneous estimation of such models however creates extensive computational burden and risks mis-identification of behavioural processes. In this context, the notion of model averaging, popular in some fields but not yet in choice modelling, presents an important avenue for improved predictions and balancing of insights across a set of candidate models.

3 Objectives

This thesis has a number of distinct objectives based on the identified research gaps above. These objectives aim to fulfil these gaps and can, broadly speaking, be classified in terms of choice modelling methodology or innovative applications. The specific aims of the work are categorised into one of these themes and are described below.

Methodological

M1: The operationalisation of decision field theory (DFT) for choice modelling applications. Our first objective is to operationalise DFT. Whilst Berkowitsch et al. (2014) developed an approach for calculating the probabilities of choosing alternatives that avoided the use of computationally-intensive simulation, their sacrifice of setting the number of preference updating steps to infinity appears not to maximise the potential of DFT as Hotaling et al. (2010) argues that DFT performs better if decision response times are short. DFT was built as a dynamic model, and we thus aim to apply it as such. We also aim to identify best practices for the implementation of DFT as well as developing it so that it can incorporate traditional extensions included in mainstream choice models, such as inter and intra-respondent heterogeneity, alternative specific coefficients and attributes and sociodemographics such as income effects. Another key task is to find a method such that a priori information about the directionality of attributes is not required.

M2: The operationalisation of the multi-attribute linear ballistic accumulator (MLBA) model. Applications of MLBA have thus far not made it beyond cognitive psychology. As a result, it has not yet been applied to large-scale choice modelling applications. Consequently, features such as alternative specific coefficients and attributes have not been implemented into a MLBA model. Like DFT, there have been numerous different versions of the model in that the underlying structure is often adapted for specific choice contexts (see for example

Cohen et al. 2017), resulting in best practices for the implementation of the model being unknown. Additionally, given that it has a number of underlying process parameters, for which the precise nature of the mathematical impacts are unknown, we aim to standardise MLBA models such that they are simpler and easier for analysts to apply in the future.

M3: The development and operationalisation of choice models based on quantum logic. Given that there has been considerable success in the transition of a quantum logic framework into cognitive psychology, this thesis aims to test whether this framework can take a further step into travel behaviour research. This thesis aims to consider different possible structures for the operationlisation of quantum logic into a discrete choice model, as well as exploring the benefits of such an approach. A key step here will be finding appropriate functions for the translation of attribute values into subjective values to be used within the quantum framework. Furthermore, we aim to test whether models with a quantum framework can provide accurate transitions between different choice contexts.

Applied

A1: To rigorously test models from mathematical psychology against mainstream choice models. The development and improvement of the models detailed above is not worth further examination if they do not yield good performances in typical choice data. We thus aim to rigorously test all the models refined in this paper against standard choice modelling techniques. This involves some of the first applications for all of these models in travel behaviour choice contexts.

A2: To test models accumulator and quantum models in the context of real world choices. Given that models in mathematical psychology have rarely been tested outside of laboratory conditions (as discussed above), we aim to develop and test these models on real data. As well as testing these models on a variety of stated preference choice datasets, we aim to test them on revealed preference data for the first time.

A3: To explore the insights generated by applications combining these models. Given the above discussion on latent class modelling and model averaging, we aim to establish the benefits of model averaging approaches in a travel behaviour context as well as testing the relative benefits of applying very different models within the different classes of a latent class model.

4 Thesis outline and contributions

This section outlines the contents of each chapter by presenting the key ideas, developments and applications that are detailed in each paper, as well as outlining how the paper addresses the research gaps identified in Section 2 and attempts to fulfil the objective(s) detailed in Section 3.

Chapter 2 presents a paper entitled 'Decision field theory: improvements to current methodology and comparisons with standard choice modelling techniques'. The focus of this paper is to take an in-depth look at the underlying mechanisms of multi-attribute decision field theory (DFT). We detail arguments in favour of using approaches such as DFT as well as discussing the previous successes and shortcomings of the model. From a methodological perspective, the paper focusses on methodological improvements which allow for a more flexible model whilst simultaneously improving the ease of estimation of the model. Additionally, the paper considers a number of important steps for increasing the behavioural flexibility of DFT, such that it can include typical features incorporated by standard choice models. For example, we detail methods for including the influence of underlying preferences towards alternatives and look at a number of methods including mixed parameters for the incorporation of both intra and inter-respondent heterogeneity. We rigorously test the increased flexibility of the model by applying it to a number of simulated datasets as well as two stated preference route choice datasets (thus providing the first application of DFT to such data). The results from the simulated datasets highlight the importance of carefully specifying a DFT model but demonstrate the improved flexibility of our new approach. The empirical results from the stated preference datasets demonstrate that DFT achieves better model fit for both estimation and forecasting than multinomial logit and random regret minimisation. We additionally discuss in detail the differences in outputs of DFT in comparison to alternative approaches.

Chapter 3 presents a paper entitled 'A careful respondent or an uncertain response: Disentangling confounding sources of increased deliberation time using decision field theory'. The aim of this paper is to provide a detailed investigation into the precise nature of the impact of the number of preference updating steps within a DFT model. A key issue with previous applications of DFT is that it has rarely been used dynamically, as simplifications have typically been made to either fix the number of preference updating steps to a large number or to estimate model outputs with simulation. However, given the computational developments detailed in Chapter 2, the probability of alternatives under DFT can now be easily analytically calculated at any time point. Whilst this was simply an estimated parameter in Chapter 2, the work in this paper focusses on uncovering the 'true' nature of DFT by considering applications on datasets with recorded choice response

9

times. We create links between response time and DFT's parameter for the number of preference updating steps, demonstrating across accommodation, route and conservation choice datasets that model fit can be improved by varying the number of DFT preference updating steps as a function of response time. Additionally, we provide some of the first comparisons of two different stuctural forms of DFT: one of which assumes that decision-makers reach an external threshold, the other for which the difference in preference between alternatives reaches some threshold. This paper also considers the confounding nature of response time, as as a longer decision time could mean that the decision-maker has considered their choice more carefully, or it could mean that they are less certain. We demonstrate how DFT can be used to disentangle these confounding sources, finding that without any measure for choice certainty, the choice response time cannot be directly proportional to the number of deliberation timesteps as faster responses can result in choices that are both more and less deterministic. Furthermore, we demonstrate that results from multinomial logit (MNL) and latent class models suggest that DFT's timestep parameter is in fact very similar to a MNL's scale parameter, suggesting that the timestep parameter simply captures how deterministic a decision-maker is from the perspective of the analyst.

Chapter 4 presents a paper entitled 'An accumulation of preference: contrasts between Decision Field Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for travel behaviour modelling'. In this paper, we provide a rigorous comparison of decision field theory (DFT) against both the multi-attribute linear ballistic accumulator model (MLBA) and traditional choice models. With MLBA having never been used before in the context of travel behaviour modelling, we comprehensively explore the underlying structure of it as well as comparing it to DFT. Crucially, with both models having never been tested in the context of route choice data (with the exception of the DFT applications in Chapter 2) we provide a detailed exploration of the impacts of normalisations for the models, given that they both have parameters that can become confounded in the context of choice only data. This paper thus serves as a guide for the considerations required when applying these models in practice. We find that both MLBA and DFT perform at least as well as multinomial logit and random regret minimisation. Results from a series of simulated datasets also support the use of both DFT and MLBA, as they demonstrate that both models can be sufficiently adapted such that parameters used in typical random utility models such as alternative specific coefficients and constants can be added effectively to DFT and MLBA. Furthermore, we develop a scale-invariant version of DFT (which was highlighted as a key future requirement in Chapter 2), meaning that a priori knowledge of the directionality of attributes is no longer required. We additionally test both DFT and MLBA on a revealed preference dataset for the first time, for which both models have better model fit than standard

choice models in both estimation and out-of-sample validation.

   Chapter 5 presents a paper entitled 'Quantum probability models: a new framework for modelling choices'. Whilst many modellers have attempted to add behavioural realism to their choice models, efforts have focussed on models that are often too similar to random utility models. This paper thus further builds on the theme that if we are to consider alternatives to RUM, we should look further afield. This paper provides the first application of models based on a quantum framework to multi-attribute, multi-alternative travel behaviour choices. Quantum probability, first developed for use in physics, has recently made a successful transition into cognitive psychology, where it has been used effectively to explain ordering and contextual effects that had previously been hard to account for with traditional cognitive accounts of behaviour. In this paper, we test whether quantum probability can take a further step into choice modelling. As well as giving a detailed explanation of how quantum probabilities can be used in choice contexts, we develop two new alternative models based on a quantum probability framework. We rigorously test both models on three route choice datasets, finding that they consistently provide better model fit than traditional choice models. Additionally, we demonstrate that models with a quantum framework can incorporate efficient transformations for adjustments in the choice context. This results in effective models for capturing the different sensitivities to attributes in best compared to worst choice, as well as capturing ordering effects of attributes and alternatives in a typical route choice dataset. Furthermore, we test whether quantum models can also be used to capture 'changing states' or equivalently 'changing perspectives' in moral contexts, finding that quantum models provide a good account of behaviour and again outperform alternative standard choice modelling approaches, with 'quantum rotations' providing effective mappings between different choice contexts.

   Chapter 6 presents a paper entitled 'New models for dynamic choice contexts: further steps towards bridging choice modelling with mathematical psychology'. In this paper, we build upon work in the previous chapters by moving accumulator models beyond the comfort of analysing experimental or stated preference choices by considering *truly* dynamic choice contexts in which the attributes of alternatives rapidly change over time. Thus far, dynamic discrete choice models have typically been used for scenarios such as driving behaviour. This is despite the fact that accumulation models such as DFT have a natural method for the incorporation of updating attributes, given that they already attempt to explain the underlying process of making a decision. This paper provides a detailed discussion of the relative merits of moving towards accumulation models for such contexts as well as providing important first steps towards the operationalisation of decision field theory in the context of changing attribute values. We then test our newly developed DFT model on a driving behaviour dataset where vehicles merge

from a on-ramp onto a motorway. We discuss in detail how DFT models can be adjusted to account for such behaviour as well as applying a number of variations of the model to the data. We additionally discuss and test the difference between analytical and simulated versions of DFT, as well as comparing them to more traditional models such as logit and probit. Results indicate that our DFT models perform comparably to alternative structures for dynamic data, with the increased flexibility which comes with a simulated DFT model demonstrating significant scope for future development. This paper also provides a first example of applying quantum models to real world data, by expanding on work in Chapter 5 to extend quantum rotation models to dynamically changing choice contexts. Our results in this Chapter also suggest that quantum models have significant scope for future development in the context of dynamic choice contexts.

Chapter 7 presents a paper entitled 'Improving forecasts and behavioural insights by applying model averaging across multiple choice models'. In this paper, we discuss and test the application of model averaging, where a single model is built by averaging across a number of candidate models. Whilst this approach is popular in, for example, weather forecasting, it has yet to become popular in mainstream travel behaviour analysis. In particular, we identify several key reasons for the application of model averaging. The first scenario is when a modeller cannot easily choose between a number of advanced models, all with some desirable properties. The second is the situation where typical advanced models cannot be used due to the size of the data and/or choice sets but where all simple models are 'unsatisfactory'. Thirdly, we demonstrate how model averaging can be used to investigate sources of heterogeneity. Through a number of empirical applications, we find that for the first two of the scenarios described above, model averaging results in a consistent improvement in model fit for both estimation and in forecasting with subsets of validation samples. Furthermore, we show that model averaging can be used to obtain elasticities and welfare measures by averaging across outputs obtained from the set of candidate models. For the third scenario, we consider two key applications comparing the results of latent class models and model averaging. The first of these looks at models where differences in attribute attendance are captured through latent class structures and the second application is in the context of decision rule heterogeneity. With this paper providing the first applications of latent class models with combinations of traditional models and both DFT and quantum models, we improve the scope for finding decision rule heterogeniety through the implementation of classes with vastly different underlying structures. Both applications discuss the contrasting results of model averaging and fully flexible latent class models, with results suggesting that latent class models may predominantly capture taste heterogeneity for individual attributes, even if the analyst seeks to uncover heterogeneity in the overall structure.

Chapter 8 presents the discussion and conclusions drawn from considering the chapters together, linking and comparing the different contributions as well as highlighting important future steps for this research.

References

Aerts, D. (2012). Quantum interference and superposition in cognition: A theory for the disjunction of concepts. In *Worldviews, science and us: Bridging knowledge and its implications for our perspectives of the world*, pages 169–211. World Scientific.

Bateman, I. J., Carson, R. T., Day, B., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R., and Wang, P. (2008). Choice set awareness and ordering effects in discrete choice experiments in discrete choice experiments. Technical report, CSERGE working paper EDM.

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.

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

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

Carlsson, F., Mørkbak, M. R., and Olsen, S. B. (2012). The first time is the hardest: A test of ordering effects in choice experiments. *Journal of Choice Modelling*, 5(2):19–37.

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

Chorus, C. G. (2015). Models of moral decision making: Literature review and research agenda for discrete choice analysis. *Journal of choice modelling*, 16:69–85.

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.

Choudhury, C. F., Ramanujam, V., and Ben-Akiva, M. E. (2009). Modeling acceleration decisions for freeway merges. *Transportation research record*, 2124(1):45–57.

References

Cirillo, C. and Xu, R. (2011). Dynamic discrete choice models for transportation. *Transport Reviews*, 31(4):473–494.

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.

Hess, S., Stathopoulos, A., and Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591.

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.

Kondyli, A. and Elefteriadou, L. (2012). Driver behavior at freeway-ramp merging areas based on instrumented vehicle observations. *Transportation Letters*, 4(3):129–142.

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.

Lipovetsky, S. (2018). Quantum paradigm of probability amplitude and complex utility in entangled discrete choice modeling. *Journal of choice modelling*, 27:62–73.

Luce, D. (1960). Individual choice behaviour. *John Wiley and Sons, New York*.

Marschak, J. (1960). Binary choice constraints and random utility indicators. *K. Arrow (Ed.), Stanford symposium on mathematical methods in the social sciences, Stanford Univ. Press, Stanford, CA.*

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

Nguyen, T. C., Robinson, J., Kaneko, S., and Nguyen, T. C. (2015). Examining ordering effects in discrete choice experiments: A case study in vietnam. *Economic Analysis and Policy*, 45:39–57.

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.

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.

Trueblood, J. and Busemeyer, J. (2010). A comparison of the belief-adjustment model and the quantum inference model as explanations of order effects in human inference. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32.

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.

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multi-attribute choice. *Psychological Review*, 117(4):1275.

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

Wang, Z., Solloway, T., Shiffrin, R. M., and Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26):9431–9436.

Yu, J. G. and Jayakrishnan, R. (2018). A quantum cognition model for bridging stated and revealed preference. *Transportation Research Part B: Methodological*, 118:263–280.

Chapter 2

Decision field theory: improvements to current methodology and comparisons with standard choice modelling techniques

Thomas Hancock[1], Stephane Hess[1], & Charisma Choudhury[1]

Abstract

There is a growing interest in the travel behaviour modelling community in using alternative methods to capture the behavioural mechanisms that drive our transport choices. The traditional method has been random utility maximisation (RUM) and recent interest has focussed on random regret minimisation (RRM), but there are many other possibilities. Decision field theory (DFT), a dynamic model popular in mathematical psychology, has recently been put forward as a rival to RUM but has not yet been investigated in detail or compared against other competing models like RRM. This paper considers arguments in favour of using DFT, reviews how it has been used in transport literature so far and provides methodological improvements to further the mechanisms behind DFT to better represent general decision making. In particular, we demonstrate how the probability of alternatives can be calculated after any number of timesteps in a DFT model. We then look at how to best operationalise DFT using simulated datasets, finding that it can cope with underlying preferences towards alternatives, can include socio-demographic variables and that it performs best when standard score normalisation is applied to the alternative attribute levels. We also present a detailed comparison of DFT and multinomial logit (MNL) models using stated preference route choice datasets and find that DFT achieves significantly better

---

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)

fit in estimation as well as forecasting. We also find that our methodological improvement provides DFT with much greater flexibility and that there are numerous approaches that can be adopted to incorporate heterogeneity within a DFT model. In particular, random parameters vastly improve the model fit.

1   An introduction to decision field theory

Random Utility Maximisation (RUM) models have dominated the field of choice modelling for over 40 years (McFadden, 2000), particularly in travel behaviour research (Ben-Akiva and Bierlaire, 1999). Recently, however, there has been increasing interest in using alternative methods to make the models flexible to accommodate departures from behaviours assumed under RUM. A key example in transport research has been random regret minimisation (Chorus, 2010; Chorus et al., 2008), which assumes that decision-makers seek to minimise the negatives rather than maximising positives. Another example comes in the form of bayesian belief networks (Parvaneh et al., 2012), which take a more heuristic approach, looking at an individual's past experiences and expectations about the different alternatives available. Whilst these new methods both make more of an effort to consider the underlying cognitive processes in decision making, another model, decision field theory (Busemeyer and Townsend, 1992, 1993), was designed purely as a cognitive model to capture the deliberation process in decision making. Decision field theory (DFT) is a stochastic-dynamic model of decision-making behaviour, which was expanded to include multi-attribute (Diederich, 1997) and then multi-alternative decision-making (Roe et al., 2001), where it was renamed multi-alternative decision field theory (MDFT)[2]. Due to the psychological roots of DFT (Busemeyer and Diederich, 2002), it has predominantly been used to explain behaviour not typically studied using 'traditional' choice models. DFT can theoretically explain similarity, attraction and compromise effects (Roe et al., 2001) and this has largely been the focus of DFT research with many papers looking into how well it can explain these context effects compared to other models (Noguchi and Stewart, 2014; Trueblood et al., 2013; Tsetsos et al., 2010). It is of course true that RUM models can also be used to test such effects, with notably nested logit being used to study the similarity effect (Guevara and Fukushi, 2016) or preference reversals (Batley and Hess, 2016). However, decision field theory further differentiates from these models by being a dynamic model. This means that it can successfully be used to study risky choices or the effect of time pressure (Busemeyer and

---

[2]Note that this version assumes that a decision-maker stops deliberating on their alternatives upon reaching some 'external' threshold after some number of preference updating steps. More generally, the model should be referred to as 'DFT'

Townsend, 1993; Diederich, 1997; Dror et al., 1999). Despite the success of DFT in explaining time and context effects, it has not often been used to explain riskless choices or decision making in general. Whilst the number of comparisons of DFT against mainstream choice models are limited, Berkowitsch et al. (2014) demonstrate that DFT performs favourably in comparison to logit and probit for consumer choice. Consequently, it is possible that DFT may be able to make the transition into mainstream choice modelling. We address this research gap in this paper by providing methodological improvements to further the mechanisms behind DFT to better represent general decision making, incorporating potential effects of socio-demographic variables and accommodating for heterogeneity. The models are rigorously compared against RUM and RRM, both for estimation and prediction, using simulated and real datasets. The remainder of this paper is organised as follows. The following section provides a comprehensive review of the multi-alternative version of decision field theory (MDFT): how it works, comparisons with other models and arguments in favour of using MDFT. Section 3 gives our methodological improvements for MDFT. Section 4 presents the data and looks at our results from using MDFT and Section 5 presents some conclusions.

2 Overview of decision field theory

Thus far, Berkowitsch et al. (2014) have provided the only comparison of MDFT against mainstream choice models. As far as we are aware, MDFT has never been compared to RRM or other alternative models from choice modelling, nor have the predictive capabilities of MDFT been tested. We do not yet know if specific types of choices will be better explained by MDFT or if certain decision-makers may be better represented by a MDFT model. In the following subsection, a summary is provided for the basic mechanisms of MDFT. We then consider arguments in support of MDFT and look further into how it has been used so far in transport research. We conclude by looking at how MDFT has been compared to RUM thus far.

2.1 Mechanisms of decision field theory

2.1.1 Basic mechanism

The main idea behind multi-alternative decision field theory is that each available alternative has a 'preference value', which updates over time[3] during a decision-maker's deliberation process for a single choice. It is assumed that

---

[3] Note that these preference values are equivalent to utilities in that they are unitless and it is the relative differences that are important.

the decision-maker continues deliberating on the alternatives available until they reach some threshold. At each step, the current values are multiplied by a 'feedback matrix' before then adding on a valence vector (which can be considered as a utility at a specific moment) at that time. In its most basic form, we have:

$$P_t = S \cdot P_{t-1} + V_t, \tag{2.1}$$

where P_t is a column matrix containing the current preference values for each alternative at preference updating step t, and S is a feedback matrix which contains three parameters (see section 2.1.3). P_{t-1} is the previous preference vector and P_0 is the initial preference vector. This is often assumed to be $[0, .., 0]'$ (Busemeyer and Diederich, 2002). Finally, V_t is the random valence vector at time t, given by:

$$V_t = C \cdot M \cdot W_t + \varepsilon_t, \tag{2.2}$$

where C is a contrast matrix, used to compare alternatives against each other, with $c_{i,i} = 1$ and $c_{i,j \neq i} = -1/(n-1)$, where n is the number of alternatives, and M is the attribute matrix containing the full set of attribute values for each alternative (which is assumed to be fully available to the decision-maker over the course of the choice process). DFT is scale-variant (Busemeyer and Diederich, 2002) and we explore the implications of failing to ensure that the attribute matrix has been appropriately scaled in section 4.3.3. At each time, t, one attribute is attended to, such that $W_t = [0..1..0]'$ with entry $j = 1$ if and only if attribute j is the attribute currently being attended to[4]. The probability of attending to attribute j is w_j. Since these weights must sum to one, a standard uniform distribution $X \sim U(0,1)$ can be used to select which attribute a decision-maker attends to at each timestep. It is assumed that there is no relationship between the timesteps, which means an attribute could be considered for several consecutive timesteps before the decision-maker considers a different attribute. 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 error is added on to allow for flexibility in the variation of probability values that MDFT predicts. The variance for the error, s, is often fixed to 1 (Trueblood et al., 2014) but can also be an estimated parameter.

---

[4]Note that this results in all but a single column of attribute matrix M (a single attribute across all alternatives) from being set to zero.

2.1.2 Calculating expected values

Expanding equation 2.1 results in:

$$P_1 = S \cdot P_0 + V_1 \tag{2.3a}$$

$$P_2 = S \cdot (S \cdot P_0 + V_1) + V_2 = S^2 \cdot P_0 + S \cdot V_1 + V_2 \tag{2.3b}$$

$$... \tag{2.3c}$$

$$P_t = \sum_{k=0}^{t-1} S^k \cdot V_{t-k} + S^t \cdot P_0 \tag{2.4}$$

The weight vectors w_j are stationary as the relative importance of the different attributes is assumed to stay constant over the course of the deliberation process. Consequently, 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 covariance matrix given by $Cov[V_t]$. Given X, a random vector, and A, a matrix of constants, we have:

$$var(A \cdot X) = A \cdot var(X) \cdot A' \tag{2.5}$$

We can now use this to calculate the expected valence. We have $E[V_t] = \mu = C \cdot M \cdot w_m$, where $w_m = [w_1, w_2, ..., w_a]'$, there are a attributes and $Cov[V_t] = \Phi = C \cdot M \cdot \Psi \cdot M' \cdot C' + s$, with $\Psi = Cov[W_t]$ and $s = Cov[\varepsilon_t]$. We can then calculate the expected value and covariance of P_t. With S being a constant, $E[P_t]$ reduces to:

$$E[P_t] = \xi_t = \sum_{k=0}^{t-1} S^k \cdot \mu + S^t \cdot P_0 \tag{2.6a}$$

$$= (I - S)^{-1}(I - S^t) \cdot \mu + S^t \cdot P_0 \tag{2.6b}$$

Roe et al. (2001) demonstrate that equation 2.5 also means that we now have:

$$Cov[P_t] = \Omega_t = Cov \left[\sum_{k=0}^{t-1} S^k \cdot V_{t-k} + S^t \cdot P_0 \right] \tag{2.7a}$$

$$= \sum_{k=0}^{t-1} \left[S^k \cdot \Phi \cdot S^{k'} \right] \tag{2.7b}$$

2.1.3 The feedback matrix

The feedback matrix allows MDFT to explain contextual effects and is defined as:

$$S = I - \phi_2 \times \exp(-\phi_1 \times D^2) \tag{2.8}$$

Where I is an identity matrix, ϕ_1 and ϕ_2 are sensitivity and memory parameters respectively, and D is some measure of distance between the attributes across alternatives. The sensitivity parameter, ϕ_1, affects how much alternatives compete with each other. This allows for the similarity effect to occur (Roe et al., 2001). The memory parameter, ϕ_2, affects the diagonal entries of the feedback matrix S. The importance of having this parameter is demonstrated by the fact that details of chosen and unchosen alternatives are often forgotten (Mather et al., 2000). A value of $s_{i,i} < 1$ indicates that preferences decays, whereas $s_{i,i} > 1$ indicates that preferences grows. Individuals have different working memory capacities (Daneman and Carpenter, 1980) and memories can grow as well as fade (Mather, 2006), an idea that appears in studies on the validity of eyewitness testimony (Christianson, 1992; Flin et al., 1992; Zaragoza and Lane, 1994). A number of different methods have been used for defining the distance, D, between alternatives in applications of DFT. Roe et al. (2001) have suggested that 'psychological' distances should be used but in application chose distances that took into account the relative position of the alternatives in the multi-attribute evaluation space. The Euclidean distance (the straight-line distance in the multi-attribute evaluation space) has also been used (Qin et al., 2013). Psychological distances can be used by including a new third parameter within the feedback matrix, w, so that distances between less competitive alternatives increase more slowly, as the Euclidean distance fails to account for the fact that some alternative attributes are more important than others (Hotaling et al., 2010). Berkowitsch et al. (2015) build on this work by creating a generalised distance function for three or more attributes.

2.1.4 Calculating probabilities

Roe et al. (2001) demonstrate that once we have results for the expected value and the covariance of preference values at time t (ξ_t and Ω_t), we can calculate the probability of choosing alternatives. They show that on the basis of the multivariate central limit theorem, P_t converges to the multivariate normal distribution. Under decision field theory, A is chosen from a set $\{A, B, C\}$ if it has a higher preference value at time t than B and C. It can therefore be

calculated[5] as

$$Pr\left[P_t\left[A\right] - P_t\left[B\right] > 0 \cap P_t\left[A\right] - P_t\left[C\right] > 0\right] = \qquad (2.9a)$$

$$\int_{X>0} exp\left[-(X-\Gamma)'\Lambda^{-1}(X-\Gamma)/2\right]/(2\pi|\Lambda|^{0.5})dX \qquad (2.9b)$$

with $X = \left[P_t\left[A\right] - P_t\left[B\right], P_t\left[A\right] - P_t\left[C\right]\right]'$, $\Gamma = L\xi_t$, $\Lambda = L\Omega_t L'$ and

$$L = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \qquad (2.10)$$

L is a matrix comprised of a column vector of 1s and a negative identity matrix of size $n-1$ where n is the number of attributes. The column vector of 1s is placed in the ith column where i is the chosen alternative. We can then use (for example) the pmnorm package in R (Genz, 1992) to calculate the probability of each alternative being chosen.

2.1.5   Simplifying the deliberation stopping process

The 'computationally dissatisfying' process of summing over powers of S (equation 2.7) can be avoided by assuming that $t \to \infty$ (Berkowitsch et al., 2014). Therefore, as long as the eigenvalues of S are less than one, $S^t \to \infty$. This reduces equation 2.6 to:

$$\xi_\infty = (I - S)^{-1} \cdot \mu \qquad (2.11)$$

More importantly, however, is the simplified form of Ω_t. From Appendix B of Berkowitsch et al. (2014), we have:

$$\overline{\Omega_\infty} = (I - Z)^{-1} \cdot \overline{\Phi} \qquad (2.12)$$

Where $\overline{\Phi}$ indicates that Φ has been transformed to a $1 \times n^2$ column vector and Z is a $n^2 \times n^2$ matrix based on S. This means that the laborious time-consuming summation in equation 2.7 can be avoided, but at the cost of assuming that all decision-makers 'take infinite' response time to make their choices.

---

[5]Note that the normal error terms within DFT, ε_t, result in the probability calculations looking similar to that of probit. However, there is additional variance generated by the random attribute attendance over multiple preference updating steps, which is captured by Ω_t.

2.2 Arguments in favour of decision field theory

There are numerous arguments in favour of using DFT. One of the main strength of DFT is that it is a dynamic model, where each alternative has a 'preference value', which fluctuates stochastically over time, as a consequence of both the random attribute attendance over the deliberation process and the random errors, ε_t. This means that DFT can explain phenomena such as preference reversal (Diederich, 1997), something that static models, such as most RUM models, cannot do. DFT is a flexible model, with two methods for a decision-maker to come to a conclusion. The decision-makers can stop deliberating either when the preference value for one alternative reaches some internal threshold value or when decision-maker reach some external factor, such as a response time limit. This is a parallel to 'satisficing' behaviour (Simon, 1957) versus maximising behaviour, a concept that was explored by Schwartz et al. (2002). Some individuals show satisficing behaviour, meaning they choose the first alternative that is deemed good enough (DFT's internal threshold), whereas others use the full time available to them to try and choose the best alternative, making a decision only when they have to (DFT's external threshold)[6]. Krosnick et al. (1996) demonstrated that satisficing behaviour can often occur when participants complete surveys and Wierzbicki (1982) provides one of the first models incorporating satisficing behaviour. It has also been demonstrated that context effects, which DFT predicts efficiently, may be fundamental to decision making (Trueblood et al., 2013), with similarity, attraction and compromise effects all appearing in a perceptual decision task. Whilst there has not yet been a large impact from neuroscience on economics (Krajbich and Dean, 2015), Busemeyer et al. (2006) suggest that the accumulation of preference, as modelled by the behaviourally derived diffusion models in DFT, closely mimics neural activations in non-human primates during perceptual decision-making tasks. For example, Gold and Shadlen (2000) found evidence of an accumulating balance of sensory information favouring one interpretation over another in the neural circuits that generate and inform a monkey's choice. Ratcliff et al. (2003) similarly found that diffusion models as opposed to Poisson models better matched the evidence accumulation process seen in neural recordings. Schall (2003) adds that it appears that there are separate neurons initiating responses- a parallel to the threshold value within DFT. DFT is also less of a 'black-box' process than typical RUM models. From a cognitive perspective, basic building blocks of cognition might be shared across a wide range of species and this bottom-top perspective is more in line with both neuroscience and evolutionary biology than the widely used top-down approach (De Waal and

---

[6]It should be noted that these are mathematically different models. Whilst the focus in this chapter is on MDFT (with external thresholds), Chapter 3 compares DFT models with internal and external thresholds.

Ferrari, 2010). DFT has a bottom-top perspective, an approach that some researchers believe to be fundamental to understanding individual's choices if we are to truly understand the underlying cognitive processes in decision making (Otter et al., 2008). To add empirical confirmation, eye-tracking data is most consistent with attribute-and-alternative-wise comparison models (Noguchi and Stewart, 2014), where comparisons are made between pairs of alternatives on single dimensions. This would suggest that DFT is an appropriate model when there are two alternatives available, although empirical confirmations for multinomial alternatives are yet to be explored.

2.3 Transport applications of decision field theory

The number of applications of DFT in transport thus far are limited and mainly theoretical. DFT has been suggested as an appropriate mechanism to explain the dynamics and high variability of choice decisions in congestion situations (Stern, 1999), due to its emphasis on an information-processing approach. Additionally, with some expansion, DFT should also be able to deal with a variety of travel situation effects including situational dynamics, type of travel, cultural habits and societal norms (Stern and Richardson, 2005). The route choice process of a daily commuter according to DFT has been conceptualised (Stern and Portugali, 1999) and DFT has also been combined with the Queuing Network-Model Human Processor to model a driver's speed control (Zhao et al., 2011). In an example of actually applying DFT in transport, it was found that given the duration to make a decision, DFT accurately predicted the percentage of participants who chose park and ride, car or bus and subway (Qin et al., 2013). While these examples demonstrate the potential of DFT in numerous important and relevant applications within transport, they all work with small scale and overly simplified hypothetical studies with limited choice scenarios. Computational limitations of DFT (Otter et al., 2008) have also limited the impact of DFT in the transport literature and there is a distinct research gap in terms of operationalising DFT for full integration in mainstream transport models. For instance, the DFT models tested so far do not account for differences in socio-demographics of the decision-makers, which have been found to have significant effects within RUM and RRM frameworks.

2.4 Decision field theory vs traditional choice models

In the only full comparison of MDFT with RUM thus far, MDFT performed as well as MNL and Probit at predicting consumer product choices made by participants (Berkowitsch et al., 2014). Additionally, when eliciting context effects, an occurrence of multiple context effects within single participants was found and MDFT then performed better than both MNL and Probit, in

part due to it being built to cope with such effects. As far as the authors are aware, MDFT has never been empirically tested against RRM. A simplified description of how a MDFT model works would be to compare it directly against a MNL or RRM model. Attributes of alternatives, M, are multiplied by W, the relative importance of the different attributes, which are equivalent to the β coefficients of MNL and RRM. We then get V, a valence vector, which can be considered as 'utility at a specific moment' and P, the total preference of alternatives vector, which is equivalent to the utility of alternatives in MNL and total regret in RRM. Whilst MDFT models do not produce utilities, we can instead use the total preference of alternatives to calculate the likelihood of alternatives (see equation 2.4 and section 2.1.4). Additionally, ϕ_1, the sensitivity parameter, allows for the similarity effect to happen under a MDFT model. This means that more similar alternatives compete more with each other, a parallel to the effect the nesting parameter, which captures the correlation across alternatives, has in a nested logit model (Daly and Zachary, 1978; McFadden, 1978; Williams, 1977)[7]. Figure 2.1 shows a connectionist interpretation of DFT (Roe et al., 2001). This demonstrates graphically how the total preference of alternatives is calculated (see equations 2.1 and 2.2 for mathematical details).
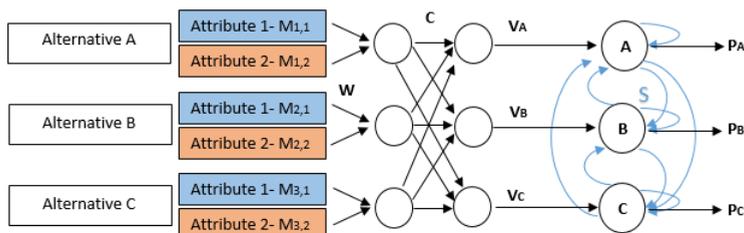
Fig. 2.1: A connectionist interpretation of DFT, adapted from Roe et al. (2001)

2.5 Summary of key DFT applications thus far

Thus far, there have not been many studies that have actually applied decision field theory (particularly within transport). Table 2.1 provides a summary of some key DFT applications, and highlights the major differences. Only one major application has been published in a transportation journal until now, with most in psychological journals such as Psychological Review. Whereas some applications only use DFT to calculate the probability of a few example alternatives, there are others that fit choice data to calculate

---

[7]Note that this is not an equivalent effect in that the similarity effect only impacts alternatives with similar attribute values in MDFT

the likelihood of multiple choices, as typically done in a choice modelling study. Very few applications estimate all parameters, with some often held constant. Decision field theory has been applied across a variety of types of choices, most often consumer choice, but also decision making in basketball. Most applications fix the number of timesteps, as prior to this paper, there was no closed form expression for calculating the probabilities of more than three alternatives at any timestep (see section 3.1). When only two alternatives are considered, the number of preference updating steps does not need to be estimated as Busemeyer et al. (2006) demonstrates that the probability of alternatives can be calculated by instead estimating an internal threshold (i.e. not using MDFT). While comparisons with RUM are limited, DFT has been compared against a number of different models including the proportional difference model (first introduced by Gonzalez-Vallejo (2002) and compared against DFT by Scheibehenne et al. 2009), the multiple linear ballistic accumulator (Trueblood et al., 2014) and traditional choice models such as logit and probit (Berkowitsch et al., 2014).

3 Improvements to decision field theory

The following section provides a method for avoiding the sacrifice by Berkowitsch et al. (2014) whilst simultaneously avoiding computationally intensive simulations. We then present methods for incorporating heterogeneity across and within decision-makers into decision field theory.

3.1 Avoiding the sacrifice of response time being set to infinity

It has been argued that the lack of analytical solutions for DFT means that it has to use computationally intensive simulations (Otter et al., 2008) and should be used with an externally controlled stopping procedure with a large value for response time (Noguchi and Stewart, 2014; Trueblood et al., 2014). However, Hotaling et al. (2010) argued that the undesirably long fixed stopping times used by Tsetsos et al. (2010) was in part why their DFT model performed worse than their own rival preference accumulation model, the leaky competing accumulator (a model designed to address challenges to previous diffusion, random walk and accumulator models), suggesting that large values for the number of preference updating steps should be avoided if possible. Berkowitsch et al. (2014) avoided arbitrarily setting the number of timesteps by fixing it to infinity, as shown in the previous section. We will now, however, show that as well as being an undesirable sacrifice, this is an unnecessary one. Firstly, we show that the following matrix can be rearranged to a more usable format as follows:

Authors	Journal	Type of estimation	Parameters	Type of choices	Key assumptions	Key DFT results
Roe et al. (2001)	Psychological Review	Probability of alternatives	Some estimated, some fixed	-	-	Demonstrated how MDFT explains context effects
Raab and Johnson (2004)	Research Quarterly for Exercise & Sport	Probability of alternatives	Some estimated, some fixed	Basketball decisions	-	Different initial preferences explained individual choices best
Scheibehenne et al. (2009)	Cognitive Science	Likelihood of multiple choices, by individual	All estimated	Monetary gambles	Uses DFT with an internal threshold	DFT performs better than the proportional difference model
Tsetsos et al. (2010)	Psychological Review	Probability of alternatives	Some estimated, some fixed		Used steady state preference states after a large number of timesteps	MDFT performs less well than LCA at explaining context effects
Hotaling et al. (2010)	Psychological Review	Probability of alternatives	Some estimated, some fixed	-	-	DFT obtains more robust predictions with internal stopping rules
Hey et al. (2010)	Journal of Risk and Uncertainty	Likelihood of multiple choices, by individual	All estimated	Monetary gambles	Uses DFT with an internal threshold	DFT predicts risky choice better than most other models considered
Qin et al. (2013)	Transportation Research Part F	Probability of alternatives	Some based on questionnaire	Mode choice	Feedback coefficients not estimated	Simulated MDFT results match survey results
Trueblood et al. (2014)	Psychological Review	Likelihood of multiple choices across decision makers	Some estimated, some fixed	Likely crime suspects	1001 timesteps	MDFT performs less well than MLBA at explaining context effects
Berkowitsch et al. (2014)	Journal of Experimental Psychology	Likelihood of multiple choices across decision makers	All estimated	Consumer decisions	Infinite time steps	MDFT performs as well as MNL and probit
Noguchi and Stewart (2014)	Cognition	Probability of alternatives	Some estimated, some fixed	Consumer decisions	1000 timesteps	Eye-tracking suggests alternatives are compared rather than individually evaluated

Table 2.1: Some key DFT applications

$$\overline{S\Phi S'} = Z\overline{\Phi} \qquad (2.13)$$

where S is the feedback matrix and Φ is the covariance of V_t as before. Again, $\overline{X}$ indicates that matrix X of size $n \times n$ has been reshaped into a column matrix of size $1 \times n^2$. Now if we start with any 3 matrices of size $n \times n$,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}$$

$$(2.14)$$

we have that for multiplying matrix A by B, the entry $[AB]_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Therefore if we set $ABC = D$, we have entries $[D]_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} [a_{il} b_{lk} c_{kj}]$. Now if we reshape D into a column matrix as before, we have $\overline{D}$ with entries:

$$\left[\overline{D}\right]_{(j-1)n+i} = \sum_{k=1}^{n} \sum_{l=1}^{n} [a_{il} b_{lk} c_{kj}] \qquad (2.15)$$

Next, we wish to create a new matrix Z of size $n^2 \times n^2$ and to reshape B into a column matrix:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n^2} \\ z_{21} & z_{22} & \dots & z_{2n^2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n^2 1} & z_{n^2 2} & \dots & z_{n^2 n^2} \end{bmatrix} \overline{B} = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \\ b_{12} \\ \vdots \\ b_{nn} \end{bmatrix} \qquad (2.16)$$

Multiplying these together gives $Z\overline{B}$ with entries $\left[Z\overline{B}\right]_i = \sum_{k=1}^{n} \sum_{l=1}^{n} \left[z_{i,(k-1)n+l} b_{lk}\right]$. This gives us

$$\left[Z\overline{B}\right]_{(j-1)n+i} = \sum_{k=1}^{n} \sum_{l=1}^{n} \left[z_{(j-1)n+i,(k-1)n+l} b_{lk}\right] \qquad (2.17)$$

Thus for $Z\overline{B} = \overline{D}$, we need only set $z_{(j-1)n+i,(k-1)n+l} = a_{il} c_{kj}$. Hence we can rearrange equation (2.13) to this more useful format by setting $A = S, B = \Phi$ and $C = S'$ and following the above steps to find the new matrix Z. We now wish to show that

$$\overline{S^n \Phi S^{n\prime}} = Z^n \overline{\Phi} \qquad (2.18)$$

To do this, We employ a proof by induction. We have that equation (2.18) holds when $n = 1$ as we know that equation (2.13) is true. This means that if we can show that equation (2.19) holds, then we will have proved that equation (2.18) holds when $n = [2, 3, 4, ...]$.

$$\overline{S^{n+1}\Phi S^{n+1\prime}} = Z^{n+1}\overline{\Phi} \tag{2.19}$$

Firstly, we set the matrices $A^n = X$, $C^n = Y$ and $Z^n = W$. Then the elements of the left side matrix of equation (2.19) are:

$$\left[A^{n+1}BC^{n+1}\right]_{ij} = [AXBCY]_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[a_{ir}\,x_{rl}\,b_{lk}\,y_{ks}\,c_{sj}\right] \tag{2.20}$$

$$\Rightarrow \overline{[AXBCY]}_{(j-1)n+i} = \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[a_{ir}\,x_{rl}\,b_{lk}\,y_{ks}\,c_{sj}\right] \tag{2.21}$$

Now for the right hand side matrix, from the previous result for equation (2.13) we can set

$$z_{(j-1)n+i,(k-1)n+l} = a_{il}\,c_{kj} \tag{2.22a}$$
$$w_{(j-1)n+i,(k-1)n+l} = x_{il}\,y_{kj} \tag{2.22b}$$

and when multiplying these matrices together, we get

$$[ZW]_{uv} = \sum_{r=1}^{n}\sum_{s=1}^{n}\left[z_{u,(s-1)n+r}\,w_{(s-1)n+r,v}\right] \tag{2.23}$$

From before we had

$$\left[Z\overline{B}\right]_{i} = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[z_{i,(k-1)n+l}b_{lk}\right] \tag{2.24}$$

so for ZW this becomes

$$\left[ZW\overline{B}\right]_{i} = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[[ZW]_{i,(k-1)n+l}\,b_{lk}\right] \tag{2.25}$$

Substituting back in equation (2.23) and we get

$$\left[ZW\overline{B}\right]_{i} = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[\sum_{r=1}^{n}\sum_{s=1}^{n}\left[z_{i,(s-1)n+r}\,w_{(s-1)n+r,(k-1)n+l}\right]b_{lk}\right] \tag{2.26}$$

Finally using equations (2.22) and rearranging, the right hand side of equation (2.19) becomes

$$\left[ZW\overline{B}\right]_{(j-1)n+i} = \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[z_{(j-1)n+i,(s-1)n+r}\,w_{(s-1)n+r,(k-1)n+l}\,b_{lk}\right]$$

(2.27a)

$$= \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[a_{ir}\,c_{sj}\,x_{rl}\,y_{ks}\,b_{lk}\right] \tag{2.27b}$$

$$= \overline{[AXBCY]}_{(j-1)n+i} \tag{2.27c}$$

Hence, we have that $Z^{n+1}\overline{B} = \overline{A^{n+1}\,B\,C^{n+1}}$ and the induction is complete. Finally, using this result, we can return to equation (2.7), which now simplifies to become

$$Cov[P_t] = \Omega_t = \sum_{k=0}^{t-1}\left[S^k \cdot \Phi \cdot S^{k'}\right] \tag{2.28a}$$

$$= \sum_{k=0}^{t-1}\left[Z^k \cdot \overline{\Phi}\right] \tag{2.28b}$$

$$= (I - Z)^{-1}\,(I - Z^t)\,\overline{\Phi} \tag{2.28c}$$

with Z being created from elements of the feedback matrix S by setting $z_{(j-1)n+i,(k-1)n+l} = s_{il}\,s_{jk}$ for $i, j, k, l \in [1:n]$. This means that the time consuming sum calculation has been removed and we can therefore return to having a finite value for t. We thus restore the original core psychological foundations of MDFT whilst simultaneously avoiding intensive calculations[8]. In section 4.2 we compare the results of different versions of MDFT, looking at the implications of making this simplification. We compare our version of MDFT (MDFT-2017), where we estimate the number of timesteps a decision-maker takes to reach a conclusion, against the previous version of MDFT (MDFT-2014), where decision-makers preferences are assumed to have stabilised over an infinite time. MDFT-2014 can be incorporated within MDFT-2017, simply by setting the number of timesteps to a high value.

3.2 Alternative specific constant (asc) parameters

One of the strengths of RUM models is their ability to measure baseline preferences of some alternatives through the use of alternative specific constants.

---

[8]Note that equation 2.28c becomes equation 2.12 for $t \to \infty$ if the eigenvalues of the feedback matrix are less than one (or equivalently $\phi_2 < 1$), in which case $S^t \to 0$ (Roe et al., 2001) and hence $Z^t \to 0$.

MDFT can partly accommodate this through P_0, the initial preference vector. Crucially, this leads to just an *initial* preference for this alternative, meaning that the preference disappears as the decision time increases and the number of timesteps becomes high. As MDFT-2014 has an infinite number of timesteps, this means that it has no method for accommodating initial preferences. However, Roe et al. (2001) used an additional weight assigned to a zero column matrix, as a way of reflecting that the decision-maker was attending to 'other irrelevant' attributes (which may actually be relevant). We can expand on this idea by having an additional attribute 'looking at other factors favouring alternative x.' This would have attribute levels of y for alternative x, and 0 for all other alternatives. We can either fix y to being a specific value, or allow it to fluctuate by adding it in as another parameter in the same way that alternative specific constants are. Depending on the number of alternatives, more of these additional attributes can be added as required. This gives us two methods for MDFT-2017 and one method for MDFT-2014 to deal with preferences towards alternatives, all of which are explored in section 4.3.2.

3.3   Adding heterogeneity

Decision field theory has almost always been implemented as a 'one size fits all' model, with an exception being Raab and Johnson (2004), who looked at individual differences in action taking within sport (although this looked at just a single DFT choice scenario, as the attributes were not clearly defined). Scheibehenne et al. (2009) and Hey et al. (2010) also considered individual differences by computing separate DFT models for each decision maker, but as far as we are aware, no studies have thus far fitted a DFT model to multiple decision makers across multiple decisions whilst simultaneously incorporating individual differences. This is surprising given that DFT has psychological origins, where individual differences tend to be better appreciated. Johnson (2006) highlighted the need for DFT to be able to explain individual differences and Liew et al. (2016) found that, as a contradiction to the findings of Berkowitsch et al. (2014), participants rarely showed all three context effects, highlighting the dangers of averaging indiscriminately and not having a method for dealing with individual differences. We believe that there is no reason that DFT cannot be expanded in exactly the same way that Multinomial Logit (MNL) has been within RUM. A 'Mixed DFT' could incorporate some of the ideas of Mixed Logit (McFadden and Train, 2000): some parameters could be changed from having a point estimate to having a distribution instead. Whilst some caution on ranges of the distributions would be required (e.g. positive only weight parameters), there is nothing to suggest that there could not be individual variation in any one the parameters within DFT. In section 4.5.3 we explore the results of using

random parameters in a MDFT model as well as the effects of using different distributions for these parameters. Furthermore, we can also consider, for example, how the income of an individual could impact the weight for attending to the costs of alternatives.

4 Empirical application

4.1 Datasets

In this section we summarise the datasets that we have used to test the explanatory and predictive power of decision field theory against other models as well as finding the best methods to maximise the performance of a MDFT model.

4.1.1 Simulated dataset A (SD-A)

The first simulated dataset contains 1,000 choice situations, each with two attributes 'A' and 'B', and two alternatives '1' and '2'. The attribute values were drawn from a uniform distribution from 1 to 10 (with values redrawn to ensure there are no dominated alternatives). After a random number, $0 < r_1 < 1$, was created for each decision, the probability of choosing alternative 1 was defined as $0.05\,W_A(A_1 - A_2) + 0.05\,W_B(B_1 - B_2) + 0.5 > r$, where W_A and W_B where the weights of the attributes, both set to 0.5 by default. We use this basic dataset with simple choices to test the ability of a MDFT model to capture the effect of underlying preferences for an alternative in section 4.3. A preference for (arbitrarily) alternative 1 is added in by defining that for any choice task with a second random number $0 < r_2 < 1$ of less than a certain value, the decision-maker would always pick alternative 1.

4.1.2 Simulated dataset B (SD-B)

The second dataset contains 8,000 choice situations, each with six attributes 'A' through to 'F' and two alternatives '1' and '2'. Each attribute value was either true or false. An MNL model was used to simulate the choices (with coefficients $\beta_A = -0.6$, $\beta_B = -0.5$, $\beta_C = -0.4$, $\beta_D = -0.3$, $\beta_E = -0.2$ and $\beta_F = -0.1$). The aim of testing this dataset is to see how well MDFT copes with binary attributes and to compare it against MNL as detailed in section 4.4.1.

4.1.3 Simulated dataset C (SD-C)

The third dataset also contains 8,000 choice situations, this time with four attributes- cost (TC), travel time (TT), number of changes (CH) and availability of seating (AS). An MNL model was again used to calculate the

probabilities of each alternative being picked (with coefficients $\beta_{TC} = -0.5$, $\beta_{TT} = -0.05$, $\beta_{CH} = -0.5$ and $\beta_{AS} = -0.5$). This time a group difference was added in, such that group '2' attached 3 times more value to β_{AS} (for instance, in real life, the decision of some travellers may be strongly affected by the availability of seating). This dataset could then be used to test the ability of MDFT to cope with socio-demographic differences as detailed in section 4.5.1.

4.1.4 Swiss stated preference dataset (SP-1)

Our first stated preference dataset comes from the Swiss value of time study (Axhausen et al., 2008), and specifically a route choice example for rail users, where 389 participants each completed 9 binary choice tasks described by 4 variables: travel time (TT), travel cost (TC), headway (HW) and the number of changes (CH).

4.1.5 UK stated preference dataset (SP-2)

The second stated preference dataset uses 10 choice tasks from each of 368 participants, all of whom are public transport commuters in the UK. Each task involves an invariant reference trip and two hypothetical alternatives. Each alternative is described by travel time, cost, rate of crowded trips, rate of delays (both out of 10 trips), the average length of delays (across delayed trips) and cost of a provision of a delay information service (Hess and Stathopoulos, 2013).

4.2 Differences between different MDFT models

In this section we compare MDFT-2014 (MDFT without a time parameter) against MDFT-2017 (MDFT with a time parameter). The MDFT model with a time parameter uses the method described in section 3.1 whilst the one without follows the method of Berkowitsch et al. (2014), where response time is set to infinity. We also compare these models against simple multinomial logit models and also two versions of random regret minimisation models (the first following the specification of Chorus 2010) and the second following (van Cranenburgh et al., 2015), incorporating μ, a parameter to estimate a profundity of regret). For SP-1, our MNL and RRM models contain five parameters, four for the attributes and one alternative specific constant. SP-2 has an additional attribute and an additional alternative, resulting in seven parameters. The μ-RRM models have six and eight parameters respectively with the addition of the μ parameter. The MDFT models have three and four parameters respectively for attributes in SP-1 and SP-2. All MDFT models additionally have sensitivity, memory and error parameters (ϕ_1, ϕ_2 and ϵ) and MDFT-2017 models also have a parameter for the number of timesteps.

Dataset	Swiss (SP-1)				UK (SP-2)			
	LL	free par.	BIC	timestep estimate	LL	free par.	BIC	timestep estimate
MNL	-1,667.97	5	3,377		-3,721.67	7	7,501	
RRM	-1,667.97	5	3,377		-3,699.49	7	7,456	
μ-RRM	-1,667.97	6	3,405		-3,698.89	8	7,463	
MDFT-2014	-1,608.65	6	3,266	∞	-3,676.34	7	7,410	∞
MDFT-2017	-1,597.30	7	3,252	10.05	-3,598.87	8	7,263	3.78

Table 2.2: Results from removing the sacrifice of setting response time to infinity

From Table 2.2 we observe that for SP-1 and SP-2 adding a time parameter results in a significant improvement in model fit (with MDFT-2017 also outperforming MDFT-2014 in out-of-sample validation, see Section 4.6). As SP-1 only has two alternatives, RRM achieves the same result as MNL. For SP-2, RRM and μ-RRM provide significantly better fit than MNL but significantly worse fit than MDFT, especially compared to MDFT-2017 (see appendix A for full MDFT model estimates). Whilst the weight estimates are similar (Table A.2), the psychological parameters also have somewhat different estimates, likely due to the fact that they have more time to have an impact with an infinite number of preference updating steps. However, these parameters have little impact on the preference values (see appendix A). For SP-1, MDFT has two more parameters than MNL, and it could be argued that BIC values do not penalise this difference enough. However, we find that the best fitting MNL model with an additional two parameters (square root terms for cost and time) has a log-likelihood of $-1,615.79$, which is still significantly worse in model fit than MDFT.

4.3   Implementation and application of decision field theory

In this section we look at how to best implement and apply a decision field theory model. We consider the implications of the weight parameters having to be greater than zero, look at methods for MDFT to incorporate underlying preferences for an alternative and look at the effect of different scaling methods being used on the attribute levels[9].

4.3.1   Implications of decision field theory weight parameters having to be greater than zero

Using the Swiss stated preference dataset (SP-1), it is quickly possible to see the effect of having undesirable attributes in MDFT. If a value for an undesirable attribute is positive and high (for example, a large cost), then an

---

[9]From here, MDFT-2017 is always used unless otherwise specified

appropriate MDFT model would factor this in by adding a negative valence to the preference value of an alternative when this attribute is considered. However, the weight parameters in a MDFT model cannot be negative, as w_i represents the proportion of time that a decision-maker looks at attribute i. This causes issues when we have 'positive', desirable attributes (such as quality), and 'negative', undesirable attributes (such as travel cost). If the attributes were to be left as they were, then due to MDFT being an accumulative model and weights being positive, there would be no way for MDFT to reflect that an alternative is more likely to be picked if an attribute level is lower. This means that MDFT will have its greatest predictive accuracy when negative attributes are ignored, and their weights are set to zero. Table 2.3 shows the log-likelihood values of SP-1 under MDFT models where some attributes are desirable/undesirable. As all four attributes are undesirable, 'negative' here means that the higher values are less desirable, whereas positive means that they have been reset such that higher, more positive values are more desirable. The table also shows the parameter estimates for the MDFT models. As MDFT has no clear starting points for estimation, we have to run a number of trials to find a suitable starting point. We set the weight attributes to be equal and use random numbers to set ϕ_1 between 0 and 10, ϕ_2 between 0 and 1, ϵ between 0 and 1,000 and t between 0 and 100. We ran 100 trials of this nature and then used the best as the starting point in the R package maxLik (Henningsen and Toomet, 2011). We found that the inclusion of the third feedback parameter, w, made an insignificant difference to the results of MDFT, therefore omitted it in these trials and used Euclidean distances in the feedback matrix. We used standard score normalisation to scale the attributes in this section, but explore scaling methods further in section 4.3.3.

Model	1		2		3		4	
MDFT LL	-2,000.61		-1,952.68		-1,724.55		-1,597.30	
MNL LL	-2,039.46		-1,976.24		-1,722.97		-1,667.97	
TT	Negative		Positive		Positive		Positive	
TC	Negative		Negative		Negative		Positive	
HW	Negative		Negative		Positive		Positive	
CH	Positive		Positive		Positive		Positive	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
w_{TT}	0.0000	0.00	0.4528	5.26	0.3214	5.86	0.3480	45.99
w_{TC}	0.3973	2.49	0.0000	0.00	0.0005	0.07	0.4691	43.20
w_{HW}	0.0001	0.00	0.0054	0.11	0.2838	15.18	0.0739	13.03
ϕ_1	0.1806	3.20	0.1206	3.08	0.5959	4.59	-0.01	-0.02
ϕ_2	0.6645	3.52	0.6748	4.83	0.6012	8.47	0.00	0.01
ϵ	11.1728	6.79	9.9875	8.28	1.4430	5.39	0.00	0.00
t	10.0038	4.00	9.0836	8.42	12.3834	5.72	10.05	12.14

Table 2.3: Parameter estimates and log-likelihoods for MDFT models for positive and negative attributes (using SP-1)

We can see from Table 2.3 that if the travel costs alone are negative

(Model 3), the parameter for travel cost, w_{TC}, quickly drops towards zero, reflecting that the MDFT model has not used the information as to do so would worsen model fit. An equivalent hindrance on a MNL model, where the beta coefficient for travel cost is fixed to zero, suffers a similar loss in log-likelihood. When headway is also negative (model 2), w_{HW} also drops to zero. The value for the error term, ϵ, has increased significantly. Hotaling et al. (2010) argue that the error term (also known as the noise variance Roe et al. 2001) should be higher for more complex tasks, meaning that higher values for ϵ would be found for less predictable decisions, as demonstrated here. Model 1, where travel time is made negative, obtains further losses in model fit. Perhaps surprisingly, the coefficient for travel cost weight increases away from zero. However, this is due to alternatives with low travel time and high cost being generally preferred to alternatives with high travel time and low cost. This is also reflected in an MNL model with just beta coefficients for travel cost and the number of changes, in which a positive value ($\beta_{TT} = 0.015$, t-ratio= 2.34) for travel cost is found. The memory parameter, ϕ_2, is higher for models with negative attributes, suggesting that MDFT predicts that the participant 'forgets' the information[10] they are looking at and that the choice is more down to chance instead, as can be seen by the higher values for ϵ. This makes sense given that when MDFT has negative attributes, preference values increase as the alternative becomes less likely to be picked, meaning that the less the attribute values are used to make the predictions, the better. If all attributes are positive, MDFT starts to outperform MNL more significantly. The memory parameter drops to zero as the accumulated preference values are now reflected in the likelihood of an alternative being chosen. In conclusion, this shows that we need to invert all negative attributes such that the higher an attribute value, the more desirable an alternative is. This will improve the performance of a MDFT model, but poses problems when we do not know if an attribute is desirable or not. If this is the case, we can simply include the attribute twice, once with the original values and once with the inverted values. The weight for one will quickly drop to zero indicating whether the attribute is positive or negative.

4.3.2   Dealing with underlying preferences

Random utility models with a multinomial logit framework deal with underlying preferences through alternative specific constants (McFadden and Train, 2000). These values directly capture market shares. MDFT has two methods for capturing shares and dealing with preferences towards an alternative. One is through the initial preference matrix P_0 and the other is by creating a new attribute favouring one of the alternatives. Here we simply

---

[10]Note that mathematically, preferences being 'forgotten' cannot be disentangled from a slower rate of preference accumulation.

add in a dummy variable for the alternatives with some fixed higher value for one of the alternatives. Table 2.4 displays the results of adding in additional MDFT parameters to deal with preferences in simulated dataset A. A parameter pr_1 indicates an initial preference for alternative 1 in P_0 and parameters w_1, w_2 indicate weights for new attributes favouring alternative 1 and 2 respectively.

Proportion always choosing alternative 1	MDFT 2017	additional w_1	additional w_1, w_2	additional pr_1		MNL
				LL	estimate	
parameters	5	6	7	6		3
0.1	-678.80	-667.18^*	-665.94	-667.26	5.69	-667.20
0.2	-686.14	-647.36	-646.13	-647.99	13.71	-647.85
0.3	-687.73	-620.14	-620.14	-621.36	18.88	-621.12
0.4	-691.64	-578.24	-577.85	-578.25	30.70	-578.14
LL of null model = -693.15 Pars = 3 (MNL), 5 (MDFT-2017)						

Table 2.4: The effect of underlying alternative preferences on MDFT models (using SD-A)

For the models in Table 2.4, we set the attribute value difference for the new parameter w_1 to 5 arbitrarily in every case. A value of -666.37 was achieved in case * when a value of 1 is used instead, suggesting that the value that is used here does not significantly impact model results. This value could be set as another parameter, but as the value has not changed significantly we have not explored this further. Whereas adding in parameter w_1 for a preference of alternative 1 makes a difference, additionally adding a weight for alternative 2 does not significantly improve model fit. This means that we can treat these parameters equivalently to alternative specific constants in random utility models in this scenario, where similarly only one parameter would be needed to capture the difference in underlying preferences between two alternatives. We can also see from Table 2.4 that adding in w_1 results in MDFT achieving similar log-likelihoods to MNL. Similar values are again achieved by adding in the parameter pr_1. As the percentage of choices where decision-makers always choose alternative 1 increases, the parameter estimate for pr_1 rises.

Using MDFT-2014, where the number of timesteps is set to infinity, results in not being able to capture biases through initial preferences. However, for case *, where 10% always choose the first alternative, MDFT-2014 obtains a log-likelihood of -667.10, indicating that biases can be captured by attention weights for specific alternatives. Without additional weights, the log-likelihood for MDFT-2014 is -679.09, in line with the MDFT-2017 result with no means to capture the bias.

4.3.3   Scaling of attributes

The most common method for scaling attributes that has been used in previous applications of DFT has been to rescale values (unity-based normalisation) to be between two values (Berkowitsch et al., 2014; Johnson, 2006). We now consider different methods for scaling the attribute values in dataset SP-1 and the effects this has on the parameter estimates for MDFT. The first method we use is unity-based normalisation, where we have a minimum value of 0 and maximum value of 1. We set $a_i = 1 - \frac{a_i - min(a)}{max(a) - min(a)}$ for each of the attributes, ensuring that we set the most desirable attribute level (lower costs and travel times) to be close to 1 and less desirable attribute levels to be close to 0. For the second method, we do not scale the attributes at all, simply setting $a_i = -a_i$. For the third method, we use standard score normalisation and set $a_i = -\frac{a_i - mean(a)}{sd(a)}$. The fourth method employs the same values as the third, with the exception that the travel time values are additionally all multiplied by 10. This allows us to test whether the relative attribute importance weights can readjust appropriately. Table 2.5 shows the weight estimates for each attribute in the different MDFT models as well as the MNL beta coefficient values for travel time (TT), travel cost (TC), headway (HW) and the number of changes made when travelling by train (CH). As with other departures from RUM, value of time and similar measures cannot be directly calculated under a MDFT model. We instead define 'relative importance (RI) of time,' as $\frac{w_{TT}/S_{TT}}{w_{TC}/S_{TC}} \times 60$, where S_i is the scale factor used for scaling attribute i, and use the value of time for the relative importance of time under MNL.

We can see from Table 2.5 that MDFT produces a higher log-likelihood value than MNL does for SP-1. This difference gradually increases as we change from scaling method 1 through to 3. It appears that because of the large range of attribute values, some of the information is lost in method 1. The high weight value for the number of changes in scale 2, 0.7757, shows that due to the lack of scaling, the decision-maker has to attend to the number of changes more often for the importance of this attribute to be reflected. The best log-likelihood value under MDFT is achieved in scale 3, suggesting that it is important to include information on means and standard deviations when scaling the attribute values before running a MDFT model. The relative importance of travel time and the relative importance of changes estimates vary depending on which type of scaling is used. In general, it appears that as the model's log-likelihood value improves, both relative importance values decrease. Table 2.5 also shows the impact of multiplying the travel times in SP-1 by 10 in MDFT scale 4. The only difference this makes for a RUM Multinomial Logit is that the travel time coefficient becomes exactly 10 times lower. As a contrast, MDFT does not equivalently have a simple change of coefficients. Instead, we see that w_{TT} has decreased from 0.3480

4. Empirical application

Model	LL value	TT	TC	HW	CH	RI of travel time CHF/ hour	RI of changes CHF/ change
MDFT scale 1	-1,640.34	0.3462 (32.53)	0.5887 (45.23)	0.0261 (12.34)	0.0390	24.43	8.88
MDFT scale 2	-1,622.83	0.0565 (17.18)	0.1390 (12.95)	0.0288 (17.83)	0.7757	24.38	5.58
MDFT scale 3	-1,597.30	0.3480 (45.99)	0.4691 (43.20)	0.0739 (13.03)	0.1090	21.35	6.48
MDFT scale 4	-1,638.43	0.0615 (24.10)	0.6202 (29.02)	0.1330 (13.76)	0.1853	28.54	8.33
MNL	-1,667.97	-0.0598 (-14.04)	-0.1318 (-9.76)	-0.0375 (-20.34)	-1.1528 (-26.56)	27.21	8.74

Table 2.5: Parameter estimates (t-ratios in brackets) for MDFT models under different types of scaling for SP-1

to 0.0615. This reflects the fact that to capture the relative importance of time, it has to be attended to less often relative to the other attributes for it to be appropriately incorporated into the model. We see a lower value of log-likelihood, with the increase in relative importance values likely being the cause. Additionally, we also calculate elasticities for a 10% increase in travel time or travel cost for the chosen alternatives, with results given in Table 2.6.

Table 2.6: Elasticities for different models for the Swiss dataset

	Time	Cost
MDFT Scale 1	-0.903	-0.827
MDFT Scale 2	-0.931	-0.881
MDFT Scale 3	-1.123	-1.127
MDFT Scale 4	-0.800	-0.659
MNL	-0.759	-0.631

The results here appear to be in line with the relative importance of travel time. MDFT scale 4 is most similar to MNL and scale 3 is the most different. Crucially, these results demonstrate that using different scaling methods will impact model performance and model outputs, thus careful consideration is required before a specification is chosen[11].

---

[11]This additionally highlights the importance of using the new scale-invariant version of MDFT developed in Chapter 4.

4.4   Differences in results between decision field theory and other models

4.4.1   Exploring the differences between RUM multinomial logit and decision field theory probabilities

The different scaling methods for the attribute values for a MDFT model has a big impact on the differences between MDFT and MNL model probability of alternatives for SP-1. Figure 2.2 demonstrates that scales 1 and 3 in particular find that when the number of changes and headway is the same for both alternatives, MDFT makes a more extreme prediction than MNL, indicated by the grey points on the figure. This does not happen under scale 2, where MDFT makes more conservative predictions.
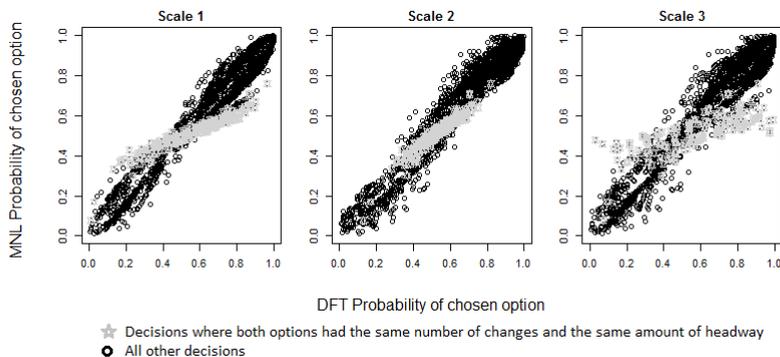
Fig. 2.2: Difference between MNL and MDFT probabilities of chosen alternatives for SP-1

Linear regression results on the absolute difference between MNL probabilities and MDFT scale 3 probabilities show that this difference is greatest when differences between alternatives are small, but that it decreases as the difference between the number of changes, travel time and headway between the alternatives increases. This is also the case in simulated dataset A, where similarly, linear regression shows that the absolute difference between MNL and MDFT probabilities decreases as the differences between the alternatives increase. Simulated dataset B finds an extremely small average difference between MNL and MDFT of $9.51e - 05$, with standard deviation 0.0041 and a largest difference of 0.008. This suggests that if alternatives only have true/false attributes, MDFT will produce very similar results to MNL. We also have from figure 2.3 that the relative importance of travel time has a significant impact on the difference between MNL and MDFT. The relative importance of travel time of an alternative is defined here as being positive if the more expensive, faster alternative is chosen and negative if the cheaper,

slower alternative is chosen. For example, a value of 50 indicates that the decision-maker is spending 50 CHF per hour saved.
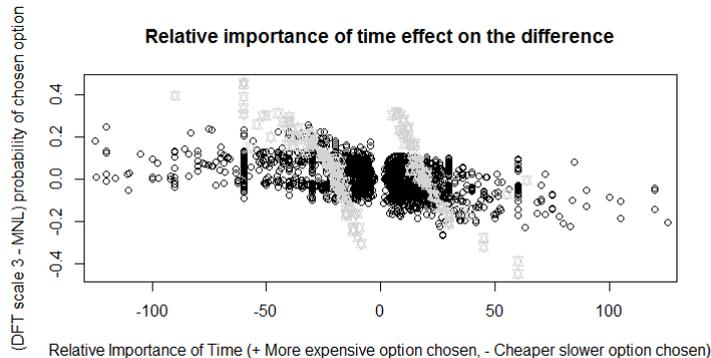
Fig. 2.3: Impact of the relative importance of time on the difference between MNL and MDFT probabilities of chosen alternatives for SP-1

From Table 2.5 we can see that MDFT (scale 3) predicts a relative importance of travel time of 21.4 whereas MNL predicts a relative importance of travel time of 27.2. This difference is reflected in figure 2.3 by the fact that the lower the relative importance of travel time is, the larger the difference between MDFT and MNL becomes in favour of MDFT. We also see for decisions that are purely a trade-off between time and cost (grey points), the impact of the value of the relative importance of time is larger for MDFT. For SP-2, it appears that MDFT gives more importance to the cost of the alternatives than MNL (see figure 2.4). Whilst both models tend to predict chosen alternatives with a probability of closer to 1 for cheaper alternatives, linear regression confirms that the cheaper the chosen alternative is relative to the unchosen alternatives, the better the fit of MDFT in comparison to MNL. Results from linear regression also imply that the reverse is true for other attributes: they have more of an impact on an MNL model. Ultimately, it appears that differences in performance between MDFT and MNL are driven by the ranges of differences for the different attributes, with increased availability of information leading to larger differences between the models.

We next consider the differences between MNL and MDFT from the perspective of individuals (see figure 2.5). Whilst individuals seem to be fairly evenly distributed for both datasets, linear regression finds that for both, MDFT tends to do better for more predictable individuals who contribute high log-likelihood values, whereas MNL provides a better fit for less predictable individuals.

UK data, individual choices

★ Decisions where the **least** expensive alternative was chosen
✕ Decisions where the **most** expensive alternative was chosen
○ All other decisions
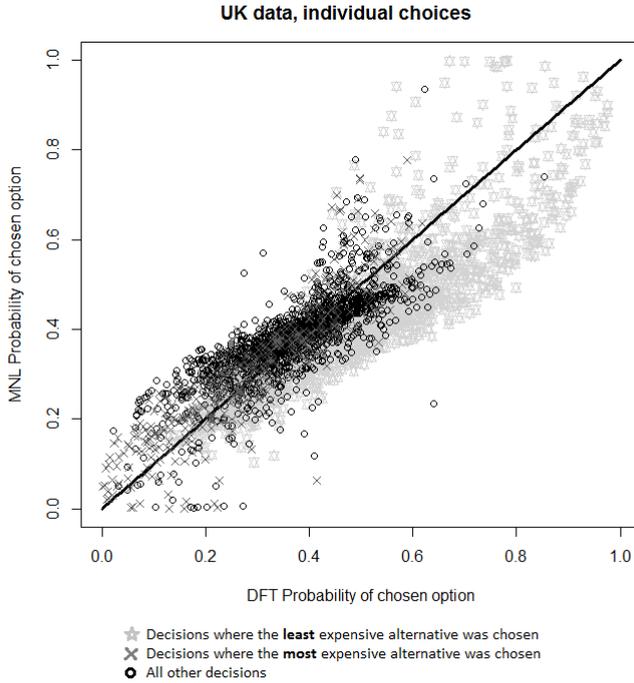
Fig. 2.4: Difference between MNL and MDFT probabilities of chosen alternatives for SP-2

4.4.2 Runtime of decision field theory

Decision field theory is a relatively slow model to run. Table 2.7 shows the runtimes for datasets SP-1 and SP-2. The runtimes are normalised relative to the runtime for MNL for SP-1 and SP-2.

Model	SP-1 Run-times (normalised)	SP-2 Run-times (normalised)
MNL	1.00	1.00
RRM	-	1.58
μ-RRM	-	2.11
MDFT-2017	**200.08**	**174.78**
MDFT-2014	**318.85**	**138.34**

Table 2.7: Relative runtimes of models for SP-1 and SP-2

Whilst MDFT-2017 takes longer to run than typical choice models, it
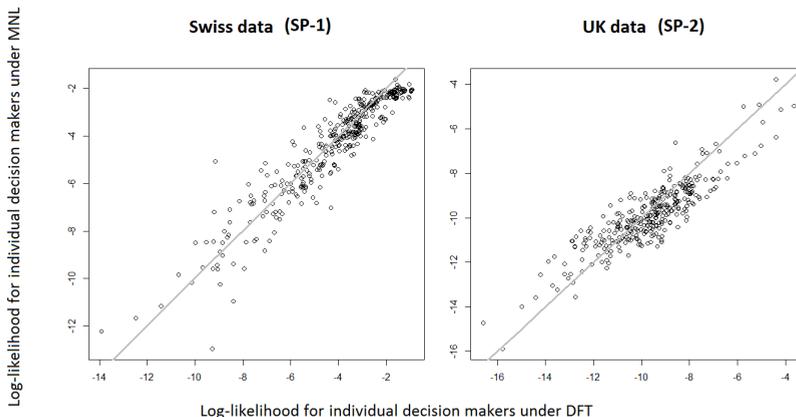
Fig. 2.5: Difference between MNL and MDFT log-likelihoods for individuals

is quicker than mixed RRM (see appendix A). Using MDFT-2014, with the number of timesteps set to infinity, reduces the runtime for SP-2 but increases it for SP-1. Runtimes for mixed decision field theory models (see 4.5.3), which are estimated by R package RSGHB (Dumont et al., 2014), vary vastly depending on the number of iterations set by the coder. A low number of iterations can be used initially to get an approximation of how well a model will work before running a more time-consuming model with more iterations.

4.5 Incorporating Heterogeneity

4.5.1 Using socio-demographic variables in decision field theory

One strength of RUM models is that they are good at using the input of socio-demographic variables to improve model accuracy. As far as we are aware, these factors have never been incorporated into MDFT. This idea is explored in this section. Firstly, we explore the impact of income on the weight parameter for travel cost, w_{TC} for SP-1. Whilst attribute parameters for MNL are independent of each other, this is not the case for MDFT weight parameters, as together they sum to 1. We hence define $w_{TC} = w_{TC} + x$, with the other parameters adjusted to $w_i = w_i - x \times \frac{w_i}{1 - w_{TC}}$, where x is defined as $x = p_{HI} \times HI$, HI is the household income and p_{HI} is a new parameter defining the strength of the impact of income[12]. For our MNL model, income is included in the utility functions by using $p_{HI} \times HI \times tc$, where tc is the travel cost of the alternative. Table 2.8 shows the results of including the income parameter on each of the standard models.

---

[12]Note that during estimation, if parameter estimates lead to weights outside the range $0 \geq w_i \geq 1$, the probability of the choices are set to extremely small values.

SP-1	Basic model	With income parameter
MNL	-1,667.97	-1,653.61
MDFT	-1,597.30	-1,592.35

SD-C	Basic model	With group parameter
MNL	-4,633.90	-4,509.65
MDFT	-4,633.63	-4,527.56

Table 2.8: Log-likelihood values for models with and without income/group difference parameter (using SP-1 and SD-C)

Whilst the improvement in model fit suggests that MDFT can capture income effects in SP-1, it is not as large as the improvement in the MNL model. This however does not appear to have a significant impact on the differences between MNL and MDFT probabilities of chosen alternatives (see figure A.1 in appendix A). We also explore the impact in a deliberately manipulated simulated dataset. Using simulated dataset C, we look at the improvements under MNL and under MDFT by including a parameter to control which group the decision belongs to. As before for MDFT, we add a factor x to the weight for seating/standing, subtracting this amount proportionally from the other weights. We set $x = p_G$ for group 1, and $x = 0$ for group 2. For MNL, we add $p_G \times S$ onto the utility for both alternatives for group 1, where S is equal to 1 or 0 depending on whether seating is available. Table 2.8 also shows the results of including a parameter for this group difference on each of the standard models. Once again, it appears that whilst MDFT improves with the inclusion of this socio-demographic variable, MNL improves more significantly.

4.5.2 Adjusting psychological parameters in decision field theory

We can additionally make small changes to the psychological parameters, ϕ_1 and ϕ_2, the sensitivity and memory measures, to attempt an improvement in fit in a MDFT model. For example, we can adjust the memory parameter depending on how many choice tasks the decision-maker has already completed. We add a new variable, ϕ_3, such that our memory parameter is now $\phi_2 + \phi_3 \times n$, where n is the task number. Alternatively, the sensitivity parameter can be similarly adjusted to be $\phi_1 + \phi_3 \times n$. Results from both of these adjustments are in Table 2.9.

Whilst the adjustments make very little difference for SP-1, there is a significant effect for SP-2, as only one parameter has been added. This suggests that there is some scope for improving the flexibility of MDFT through more complex feedback matrix structures.

Model	Swiss (SP-1)	UK (SP-2)
MDFT	-1,597.30	-3,598.87
ϕ_1 adjusted	-1,597.04	-3,592.15
ϕ_2 adjusted	-1,597.01	-3,594.78

Table 2.9: Log-likelihood values for models with and without adjustments to the memory and sensitivity parameters

4.5.3 Adding heterogeneity: Mixed MDFT

Our final effort to add heterogeneity to a MDFT model is to use random parameters. For both MDFT with and without a parameter for the number of timesteps, a significant improvement in fit is found (Table 2.10). For weight parameters, which cannot be less than zero, we use truncated normal distributions. We trial both normal and truncated normal distributions for the remaining parameters. These are then compared against MDFT models with fixed parameters as well as against a MDFT model without a time parameter with truncated normal distributions for all parameters. All mixed models are estimated using Dumont et al. (2014)'s R package 'RSGHB'.

Data	Model	Time	weights	other pars	parameters	LL	BIC
Swiss	1	yes	fixed	fixed	7	-1595.85	3248.826
Swiss	2	no	fixed	fixed	6	-1595.88	3240.725
Swiss	3	yes	truncated normal	normal	14	-1450.39	3015.031
Swiss	4	yes	truncated normal	truncated normal	14	-1438.39	2991.031
Swiss	5	no	truncated normal	truncated normal	12	-1430.41	2958.75
UK	1	yes	fixed	fixed	8	-3598.87	7263.425
UK	2	no	fixed	fixed	7	-3676.34	7410.155
UK	3	yes	truncated normal	normal	16	-3156.27	6443.911
UK	4	yes	truncated normal	truncated normal	16	-3140.09	6411.551
UK	5	no	truncated normal	truncated normal	14	-3190.23	6495.409

Table 2.10: Log-likelihoods of mixed decision field theory Models

For both datasets, vast gains are made by using random parameters. A better fit is found if truncated normal distributions are used for all parameters rather than just the weight parameters. Using random parameters in Berkowitsch et al. (2014)'s version of MDFT results in a lower BIC value for the SP-1 but a much higher one for SP-2. Whilst we do not run mixed multinomial logit or mixed random regret models, Hess et al. (2016) do run these models on the same UK dataset (see appendix A for a table of results). They find that their best fitting model is a mixed μ-RRM model, which achieves a log-likelihood of $-3,174.96$ with a BIC of $6,456.66$. Whilst Mixed MDFT has a better fit here, more rigorous trials and replications would be required to test the models against each other fairly.

4.6 Predictive capabilities of decision field theory

As discussed in section 1, previous researchers have only compared the goodness-of-fit of MDFT, and its performance in the context of forecasting has not been tested before. We have looked at the predictive capabilities of MDFT (both versions) on both of our route choice stated preference datasets, SP-1 and SP-2. We adopt the method used by Frejinger and Bierlaire (2007), using 80% subsets of the data for estimation and the remaining 20% for validation. We split individuals into five equally sized subsets, which are used in turn as validation subsets. We fit a 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%. The results for SP-1 and SP-2 are displayed in Tables 2.11 and 2.12, respectively.

For both datasets, MDFT-2017 tends to outperform MDFT-2014 in estimation. For the validation subsets, the models perform similarly for the Swiss dataset, but MDFT-2017 outperforms MDFT-2014 for the UK dataset. MDFT-2017 additionally outperforms MNL in all models across both datasets with the exception of the final validation subset for the Swiss dataset. It has better model fit than μ-RRM in all but one UK validation subset.

Model	Swiss (SP-1)		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MNL	Estimated 80%	Log-likelihood	-1,361.54	-1,322.15	-1,333.61	-1,316.08	-1,334.78
		BIC	2,762.76	2,683.98	2,706.90	2,671.85	2,709.27
5 parameters	Forecasted 20%	Log-likelihood	-307.79	-347.04	-335.58	-352.27	-333.65
		BIC	648.35	726.85	703.93	737.31	700.00
MDFT-2017	Estimated 80%	Log-likelihood	-1,307.82	-1,262.46	-1,286.56	-1,265.74	-1,263.26
		BIC	2,671.19	2,580.48	2,628.69	2,587.05	2,582.10
7 parameters	Forecasted 20%	Log-likelihood	-290.46	-336.32	-311.31	-332.01	-334.73
		BIC	626.80	718.52	668.49	709.90	715.24
MDFT-2014	Estimated 80%	Log-likelihood	-1,316.91	-1,273.26	-1,286.98	-1,274.75	-1,275.00
		BIC	2,681.43	2,594.14	2,621.58	2,597.12	2,597.63
6 parameters	Forecasted 20%	Log-likelihood	-293.24	-337.81	-322.68	-335.01	-334.32
		BIC	625.81	714.94	684.68	709.34	707.88

Table 2.11: Log-likelihoods for the estimated and forecasted datasets for MNL, MDFT-2014 and MDFT-2017 (using SP-1)

Model		UK (SP-2)		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MNL	Estimated 80%	Log-likelihood		-2920.64	-2972.21	-3004.15	-3009.39	-2970.73
		BIC		5,897.18	6,000.33	6,064.20	6,074.70	5,997.38
7 parameters	Forecasted 20%	Log-likelihood		-807.94	-750.12	-719.62	-713.78	-751.99
		BIC		1,662.13	1,546.49	1,485.49	1,473.71	1,550.13
RRM	Estimated 80%	Log-likelihood		-2904.35	-2958.92	-2986.85	-2993.27	-2944.96
		BIC		5,864.61	5,973.74	6,029.60	6,042.47	5,945.84
7 parameters	Forecasted 20%	Log-likelihood		-801.39	-741.45	-714.52	-707.74	-756.03
		BIC		1,649.02	1,529.14	1,475.29	1,461.64	1,558.22
μ-RRM	Estimated 80%	Log-likelihood		-2904.04	-2958.25	-2986.44	-2993.16	-2944.01
		BIC		5,871.97	5,980.40	6,036.77	6,050.24	5,951.93
8 parameters	Forecasted 20%	Log-likelihood		-801.18	-741.46	-714.37	-707.43	-756.27
		BIC		1,655.21	1,535.77	1,481.60	1,467.60	1,565.29
MDFT-2017	Estimated 80%	Log-likelihood		-2854.43	-2873.21	-2883.98	-2919.32	-2854.84
		BIC		5,772.74	5,810.32	5,831.84	5,902.56	5,773.60
8 parameters	Forecasted 20%	Log-likelihood		-747.82	-726.67	-717.01	-681.45	-747.14
		BIC		1,548.48	1,506.19	1,486.87	1,415.64	1,547.02
MDFT-2014	Estimated 80%	Log-likelihood		-2895.03	-2938.13	-2958.51	-2971.81	-2934.58
		BIC		5,845.96	5,932.15	5,972.93	5,999.55	5,925.09
7 parameters	Forecasted 20%	Log-likelihood		-783.87	-738.63	-728.17	-705.38	-744.14
		BIC		1,613.98	1,523.51	1,502.59	1,456.90	1,534.43

Table 2.12: Log-likelihoods for the estimated and forecasted datasets for MNL, MDFT, RRM and μ-RRM (using SP-2)

48

5   Conclusions

This paper provides methodological improvements to further the mechanisms behind MDFT to better represent general decision making, as well as rigorously comparing MDFT against traditional choice models. We also consider multiple mechanisms for incorporating heterogeneity within and across decision-makers within a MDFT model. Prior to our work, there was one comparison between MDFT and mainstream choice models (Berkowitsch et al., 2014). In this paper, we provide further evidence that MDFT can be a competitive rival to traditional choice models. Perhaps most significantly, MDFT achieves a better model fit than an MNL model in both of our stated preference datasets. MDFT also outperforms RRM in SP-2. We demonstrate, for the first time, that MDFT can step away from being a 'one-size-fits-all' model and incorporate heterogeneity in a number of different approaches. Whilst only small gains are made with the incorporation of socio-demographic variables into the weight parameters, a vastly significant gain is found with 'Mixed MDFT', where the parameters are random with either normal or truncated normal distributions. Large gains are also made with the inclusion of additional weights or parameters to deal with underlying preferences towards alternatives. Whilst we have made a brief start on the inclusion of socio-demographic variables, future work could explore this much further. For example, it could be possible that income effects are in part captured by the deliberation process in a MDFT model and tests could be done to see if there is a relationship between income and any of the psychological parameters in MDFT[13]. Additionally, the importance of the psychological parameters needs to be tested. It remains to be seen whether the sensitivity parameter is as efficient at capturing correlation across alternatives as the structural parameter in a nested logit model. It could also be that the psychological parameters are more important in risky choice, but the weight parameters are more important in riskless choice. As we have only tested riskless choice datasets here it could be that socio-demographic effects are easier to find in risky choice. Age, gender and personality have all been found to have an impact on risk-taking behaviour (Harris et al., 2006; Lauriola and Levin, 2001; Mata et al., 2011) and therefore it could be possible that some of these effects are captured in MDFT's psychological parameters. Section 4.4.1 suggests that the differences between using MNL and MDFT to explain an individual's decisions are not vast. However, it would appear that MDFT provides slightly more extreme predictions, with more predictable individuals being better explained by MDFT in comparison to MNL, and the converse being true for more random individuals. This is perhaps surprising given MDFT was originally used mostly for risky choices. Future work could look

---

[13]We thank an anonymous reviewer for this suggestion.

at whether MDFT differentiates more or less than traditional choice models on different kinds of datasets. MDFT could also easily be incorporated into one or more of the classes in a latent class structure and thus we could see if individual decision makers are better explained by MDFT or another model directly. This paper provides a method for calculating the probability of alternatives under a decision field theory model whilst simultaneously avoiding computationally intensive simulation and not setting the decision time for decision-makers to infinity. Using this method provides a number of benefits compared to using Berkowitsch et al. (2014)'s method. Our new method provides a better fit for the two stated preference datasets we apply it on and it provides a vastly greater amount of flexibility. Firstly, the response time to make a decision can now be simply incorporated into the model: the number of timesteps could vary proportionally to the time taken by the decision-maker[14]. Secondly, the memory parameter can now be negative, reflecting that preferences can inflate as well as deteriorate over time (Mather et al., 2000). Finally, and perhaps most crucially, initial preferences can play an important role in our version of MDFT. This means that, for example, MDFT should easily be able to explain a status quo bias. In conclusion, it appears that the restoration of a time parameter in MDFT results in a far more realistic psychological model for understanding choice behaviour. The implementation of decision field theory does not come easily. A standard MDFT model takes up to 200 times longer to run than a Multinomial Logit model. Future efforts should look at reducing this runtime as well as removing the scale-variant nature of MDFT, as currently we need to know if an attribute is desirable or not before we can incorporate the importance of the attribute into a MDFT model. We have shown that a MDFT model can be specified to include underlying preferences through the initial preference matrices, while additional weight parameters can be included to capture socio-demographic effects, for example. The outputs from the model provide rich insights into behaviour, but it is clear that traditional measures such as the value of time cannot be obtained from a MDFT model. This, however, is typical for departures from RUM, with Dekker (2014) highlighting the difficulties of using value of time measures obtained from random regret minimisation. As is the case with any other departures from RUM, a user thus needs to carefully make a decision of whether the increased behavioural richness of the model and the improvement in fit (both in estimation and in forecasting) is more important than an ability to produce measures for welfare analysis. In addition, the underlying psychological foundations of MDFT may be more suited than a purely microeconomic model at incorporating the increasing number of behavioural and processing indicators that are becoming available. For

---

[14]See also Chapter 3 for examples of how response time can be incorporated into a MDFT model.

example, electroencephalogram (EEG) recordings and eye-tracking have already been used to understand and predict choices (Khushaba et al., 2013; Telpaz et al., 2015; Uggeldahl et al., 2016). Given that we have relaxed the assumption on the number of timesteps for a MDFT model, we could now test MDFT on a dynamic revealed preference dataset where the attribute levels of an alternative change over time. Response times, EEG, deliberation times or eye-tracking information could be incorporated into a MDFT model, to lessen the requirement of estimation for the number of deliberation timesteps at each point as the choice set changes. Testing MDFT on such a dataset would enable us to look at the validity of underlying behavioural assumptions of a MDFT model. Additional information could also potentially be used to determine whether a decision-maker has come to an internal or external threshold under a MDFT model when making a decision. This would allow for a MDFT model to predict a decision-maker's level of confidence or uncertainty in their choice. For example, eye-tracking information showing which attributes are considered last could inform how likely a decision-maker is to come to a conclusion through satisficing, when an alternative reaches a certain preference value. Overall, this means that there is much need for further research into MDFT, which with its good results for both estimation and forecasting, appears to otherwise be a promising future model for the choice modelling community.

Acknowledgements

References

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J., and Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New Swiss results. *Transport Policy*, 15(3):173–185.

Batley, R. and Hess, S. (2016). Testing for regularity and stochastic transitivity using the structural parameter of nested logit. *Transportation Research Part B: Methodological*, 93:355–376.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, pages 5–33. Springer.

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.

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.

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

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.

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.

Christianson, S.-Å. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological bulletin*, 112(2):284.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. *Determinants of travel choice*, 335:357.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4):450–466.

De Waal, F. B. and Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in cognitive sciences*, 14(5):201–207.

Dekker, T. (2014). Indifference based value of time measures for random regret minimisation models. *Journal of choice modelling*, 12:10–20.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41(3):260–274.

References

Dror, I. E., Basola, B., and Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & cognition*, 27(4):713–725.

Dumont, J., Keller, J., Carpenter, C., and Dumont, M. J. (2014). Package 'rsghb'.

Flin, R., Boon, J., Knox, A., and Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, 83(3):323–336.

Frejinger, E. and Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, 41(3):363–378.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1(2):141–149.

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

Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109(1):137.

Guevara, C. A. and Fukushi, M. (2016). Modeling the decoy effect with context-rum models: Diagrammatic analysis and empirical evidence from route choice sp and mode choice rp case studies. *Transportation Research Part B: Methodological*, 93:318–337.

Harris, C. R., Jenkins, M., and Glaser, D. (2006). Gender differences in risk assessment: Why do women take fewer risks than men? *Judgment and Decision Making*, 1(1):48.

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

Hess, S., Beck, M., and Crastes dit Sourd, R. (2016). Can a better model specification avoid the need to move away from random utility maximisation? Transportation Research Board (TRB) 96th Annual Meeting.

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.

Hey, J. D., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, 41(2):81–111.

References

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.

Johnson, J. G. (2006). Cognitive modeling of decision making in sports. *Psychology of Sport and Exercise*, 7(6):631–652.

Khushaba, R. N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B. E., and Townsend, C. (2013). Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (eeg) and eye tracking. *Expert Systems with Applications*, 40(9):3803–3812.

Krajbich, I. and Dean, M. (2015). How can neuroscience inform economics? *Current Opinion in Behavioral Sciences*, 5:51–57.

Krosnick, J. A., Narayan, S., and Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation*, 1996(70):29–44.

Lauriola, M. and Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, 31(2):215–226.

Liew, S. X., Howe, P. D., and Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic bulletin & review*, pages 1–8.

Mata, R., Josef, A. K., Samanez-Larkin, G. R., and Hertwig, R. (2011). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences*, 1235(1):18–29.

Mather, M. (2006). Why memories may become more positive as people age. *Memory and emotion: Interdisciplinary perspectives*, pages 135–158.

Mather, M., Shafir, E., and Johnson, M. K. (2000). Misremembrance of options past: Source monitoring and choice. *Psychological Science*, 11(2):132–138.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. (2000). Disaggregate behavioral travel demand's rum side. *Travel Behaviour Research*, pages 17–63.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, pages 447–470.

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.

References

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.

Parvaneh, Z., Arentze, T., and Timmermans, H. (2012). Understanding travelers' behavior in provision of travel information: a bayesian belief approach. *Procedia-Social and Behavioral Sciences*, 54:251–260.

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.

Raab, M. and Johnson, J. G. (2004). Individual differences of action orientation for risk taking in sports. *Research quarterly for exercise and sport*, 75(3):326–336.

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 connectionst model of decision making. *Psychological review*, 108(2):370.

Schall, J. D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Current opinion in neurobiology*, 13(2):182–186.

Scheibehenne, B., Rieskamp, J., and González-Vallejo, C. (2009). Cognitive models of choice: Comparing decision field theory to the proportional difference model. *Cognitive science*, 33(5):911–939.

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.

Simon, H. A. (1957). Models of man; social and rational.

Stern, E. (1999). Reactions to congestion under time pressure. *Transportation Research Part C: Emerging Technologies*, 7(2):75–90.

Stern, E. and Portugali, J. (1999). Environmental cognition and decision making in urban navigation. *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes. John Hopkins Press, Baltimore, MD*, pages 99–118.

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

Telpaz, A., Webb, R., and Levy, D. J. (2015). Using eeg to predict consumers' future choices. *Journal of Marketing Research*, 52(4):511–529.

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.

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

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multi-attribute choice. *Psychological review*, 117(4):1275.

Uggeldahl, K., Jacobsen, C., Lundhede, T. H., and Olsen, S. B. (2016). Choice certainty in discrete choice experiments: Will eye tracking provide useful measures? *Journal of choice modelling*, 20:35–48.

van Cranenburgh, S., Guevara, C. A., and Chorus, C. G. (2015). New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice*, 74:91–109.

Wierzbicki, A. P. (1982). A mathematical basis for satisficing decision making. *Mathematical modelling*, 3(5):391–405.

Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.

Zaragoza, M. S. and Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):934.

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.

Chapter 3

A careful respondent or an uncertain response: Disentangling confounding sources of increased deliberation time using decision field theory

Thomas Hancock[1], Stephane Hess[1], & Charisma Choudhury[1]

Abstract

Decision field theory (DFT), although popular in mathematical psychology, has only recently been used in choice modelling for consumer and travel choices. A key difference that DFT has from standard choice models is that it has preference values for each alternative that update over time. This results in a different probability of picking each alternative depending on how long a decision-maker considers their alternatives. However, the computational complexities of DFT have resulted in failures to utilise its dynamic nature. Recent advances in the underlying computational methods for DFT have allowed for the analytical calculation of the probability of alternatives at any time point. Consequently, the number of preference accumulation steps can be estimated as a function of choice response time. We demonstrate that the model fit for DFT models can be improved by considering response times across route, accommodation and conservation programme choice contexts. We also explore the confounding nature of choice response time, with a key assumption within DFT and other accumulation models being that preference grows over time, contradicting a well-known result that a longer response time often indicates a less certain and hence less deterministic choice from a decision-maker. In line with DFT and preference accumulation, we find

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)

that across all three datasets, a longer mean response time indicates that a decision-maker is more deterministic. However, within a decision-maker, our models suggest that fast decisions are typically more deterministic, demonstrating that a longer response time indicates a less certain decision. Furthermore, results from multinomial logit (MNL) models suggest that DFT's time parameters performs similarly to a MNL's scale parameter. This suggests that without further consideration of measures such as choice certainty, DFT may struggle to truly capture the process of preference updating, as the time parameters are not necessarily correlated with response time and instead simply capture how deterministic a choice is.

1 Introduction

Decision field theory (DFT), first developed in the 1990s (Busemeyer and Townsend, 1992, 1993) is a dynamic, stochastic choice model in which the preference for each alternative updates over time, changing the probability with which each alternative is chosen. These preference values update at each accumulation step as the decision-maker considers the different attributes of the alternatives. The decision-maker then comes to a conclusion either when the preference value for an alternative reaches some satisfactory internal threshold value (equivalent to satisficing (Kaufman, 1990; Schwartz et al., 2002), where a participant chooses one of the alternatives if it is 'good enough') or when the decision-maker reaches some external threshold (such as running out of time). For example, a voter may not have a firm preference for a candidate but be forced to make a decision on election day. At this point the alternative with the highest preference value is chosen.

An analyst needs to decide whether to use DFT with internal or external thresholds. DFT was initially used as a model for understanding risky choice decisions, with internal thresholds used if the decision-maker chose when to stop deliberating, and external thresholds used if a time restriction was imposed. Both versions were developed for two alternatives with two attributes but DFT with an external threshold has since been expanded to allow for multiple attributes and multiple alternatives, and renamed multialternative decision field theory (MDFT, Roe et al. 2001). Decision field theory in various forms has been used widely across the mathematical psychology literature, having been used to model a variety of choices including monetary gambles (Schall, 2003), decision-making in sport (Raab and Johnson, 2004), likely crime suspects (Trueblood et al., 2014) and consumer decisions (Noguchi and Stewart, 2014). However, it has only recently been compared to models developed in econometric choice modelling (Berkowitsch et al., 2014). This is in part due to the complexity of calculating the probability of alternatives being chosen under a DFT model. In particular, for a DFT model with internal

thresholds, simulation is often required as there is no closed-form solution for the probability for which each alternative is chosen if there are more than two alternatives. Simulation can also be used for MDFT models with an external threshold (Turner et al., 2018). Many applications thus far have chosen to avoid this computationally-intensive procedure by using MDFT and fixing the number of preference accumulation steps to a high value (Berkowitsch et al., 2014; Cohen et al., 2017; Trueblood et al., 2014; Tsetsos et al., 2010). Alternatively, fixing the number of preference accumulation steps to infinity results in a simpler closed-form analytical solution for the probability of alternatives (Berkowitsch et al., 2014). However, crucially, both these approaches lose MDFT's dynamic nature. Thus, there has been a requirement to improve the computational methods behind MDFT such that the number of preference accumulation steps has a better behavioural underpinning and does not need to be set to an arbitrary value. In Chapter 2, we propose computational developments meaning that the probability with which each alternative is chosen can in fact be calculated simply after any number of preference accumulation steps in a MDFT model. The important question that remains is what these preference accumulation steps actually represent mathematically in a choice process. The work in this chapter considers how this feature is linked to choice response time.

There have been a number of studies considering choice response time in choice modelling, often with different foci and aims. For example, consistent differences in response times are found depending on the size of the difference between travel times and costs in a choice task compared to those from a reference trip (Börjesson and Fosgerau, 2015), with increased response times for larger travel time differences and decreased response times for larger differences in costs. There have also been suggestions that choice response times reflect how much cognitive effort a participant uses (Rose and Black, 2006), in which case longer response times would suggest more deterministic behaviour. Whilst Qin et al. (2013) demonstrated that the probability of each alternative at each timepoint under a MDFT model could be matched with the proportion with which each alternative was chosen under different time restrictions, this current paper demonstrates for the first time how the decision-maker's response time can be naturally incorporated into a DFT model such that the probability of alternatives being chosen across multiple choices is impacted. A key issue with such an approach is in accounting for the different reasons a choice might take more time. Whilst a complete model would control for such features, factors such as measurement errors (i.e. whether a recorded response time accurately reflects when the decision-maker came to a conclusion), levels of concentration of the decision-maker and choice certainty are all difficult to account for. Whilst the work in this paper does not control for such considerations, we demonstrate how an analyst might begin to restore the dynamic nature of DFT, through the inclusion

of response time.

Thus far, it has been typical for analysts to use decision field theory with an internal threshold if the decision-maker is free to choose when to conclude deliberating on their choice and decision field theory with an external threshold if a time limit is imposed. However, we argue that the choice of which version to use is not so simple, with complications for both versions often not accounted for. For example, a strict external threshold does not account for the different speeds at which the different decision-makers may accumulate evidence, nor any differences in method. A certain number of iterations may result in weak evidence in favour of an alternative for some individuals, but far greater evidence for that same alternative for others. Additionally, internal thresholds cannot be precisely measured and are likely to vary vastly both across and within decision-makers. Furthermore, a decision-maker may not choose an alternative because it satisfices, but because they do not wish to deliberate on a choice any longer. This could either be viewed (analytically) as a decision-maker reaching a self-imposed time threshold, or an alternative reaching a lowered preference evidence threshold.

However, regardless of whether the decision-maker stopped due to an internal or external threshold, we know that under DFT model assumptions, the alternative chosen has the highest preference value at the moment when the decision-maker concludes the deliberation process. Whilst we cannot know how many iterations of preference value updating have occurred, it is possible to estimate the number of iterations as a function of the choice response time, which can be recorded. As both versions of DFT have a parameter related to the number of iterations of preference updating, we can thus restore DFT to being a properly dynamic model in which the time taken to make a decision impacts the probability of each alternative being chosen. This paper considers state-of-the-art implementations of both versions of DFT in detail. We demonstrate a number of methods for incorporating the response time in the models, thus beginning the process of restoring DFT to being a properly dynamic model in which the amount of time taken to make a choice influences the probability with which each alternative is chosen. We then provide a number of empirical applications, demonstrating the increased flexibility of having free rather than fixed time parameters, as well as detailing the results of the different methods for including response times in the DFT models. These different models as well as comparisons with multinomial logit models allow us to investigate the nature and precise function of the time parameters in DFT models.

The remainder of this paper is organised as follows. First, we present the methodology behind decision field theory with internal and external thresholds, demonstrating how the probability with which each alternative is chosen can be calculated. Next, we demonstrate how response time can be incorporated into the DFT models. We then give an outline of the latent class

and multinomial logit models that are used in this paper. We next detail our empirical applications, where we test the impact of including response time for models on three very different stated choice datasets. Finally, we finish with some conclusions and present directions for future research.

2 Methodology

In this section we first describe how the probability with which each alternative is chosen can be calculated in decision field theory models, both with external (time) thresholds and internal (evidence) thresholds. We then demonstrate how reparameterising the time parameter results in a natural method for choice response times to be incorporated into the model. Finally, we briefly introduce the latent class and multinomial logit models that are used in the empirical applications of this paper.

2.1 Decision field theory with an external threshold (MDFT)

For a full description of the theory and estimation of MDFT models, readers should refer to Section 2 and 3 of Chapter 2.

2.2 Decision field theory with an internal threshold

2.2.1 Basic theory

For our decision field theory models using an internal threshold (which we henceforth call DFT-I), we consider the original specification of DFT by Busemeyer and Townsend (1992, 1993), which considers two alternatives, A_R and A_L. This differs from MDFT in a few small but distinct ways. Similarly to MDFT, the new preference values are a function of previous values and a number of other parameters. As DFT-I is only analytically solvable for two alternatives, we here only consider binary choice scenarios[2]. This means that we can consider just a single value, $P_\tau = Pref_\tau(A_R) - Pref_\tau(A_L)$. Alternative A_R is chosen when P_τ reaches some internal threshold θ, whereas A_L is chosen if P_τ reaches $-\theta$. The underlying key assumptions are still the same in that at each step, a single attribute k is attended to with probability w_k. This means that there is still a mean valence input, denoted δ, dependent on the valences v_R and v_L, which are weighted sums of the respective attributes for A_R and A_L (hence relative importance weights, w_k are still required for

---

[2]For the full equations for DFT using an internal threshold for more than two alternatives, readers should refer to the Appendix of Busemeyer and Townsend (1993). Further work should also consider searching for analytical solutions for more than two alternatives, as this is currently a major limitation for implementations of DFT-I.

each attribute). The preference difference between the two alternatives then updates according to:

$$P_t = [1 - (s + c) \cdot h] \cdot P(t - h) + [\delta \cdot h + \epsilon(t)], \tag{3.1}$$

where s is a growth-decay parameter, which, similarly to ϕ_2 for MDFT, effects whether recently considered attributes or initially considered attributes have more impact. The parameter c is a goal-gradient parameter, which is used to explain the effect that 'avoidance-avoidance' decisions (choosing between two negatives) take longer than 'approach-approach' decisions (choosing between two positives) and is specified precisely in Equation 3.5. δ is the mean valence input, h is a time unit and $\epsilon(t)$ is the error input, equivalently drawn from a normal distribution with mean zero, and variance $h \cdot \sigma^2$. Crucially, as a contrast to MDFT, the error variance is not assumed to be uncorrelated with the mean valence input. Instead, the variance of the error is defined directly as the input variance (Busemeyer and Townsend, 1993):

$$\sigma^2 = Var[V_R - V_L] = \sigma_R^2 + \sigma_L^2 - 2 \cdot \sigma_{RL}, \tag{3.2}$$

with σ_R^2 the variance of the valence for choice alternative A_R, σ_L^2 the variance of the valence for choice alternative A_L, and σ_{RL} the covariance between the two.

We next detail the precise specification for the various parameters discussed above. First, to control for underlying biases towards an alternative, the initial value, P_0 is set to some value z, which is typically set as an increasing function of the mean valence input δ and the internal threshold θ. We follow Busemeyer and Townsend (1993)'s definition[3]:

$$z = tanh(z^* \cdot \delta) \cdot \theta, \tag{3.3}$$

where z^* is a parameter to be estimated. Next, the internal threshold, θ, is defined to control speed-accuracy trade-offs (Busemeyer and Townsend, 1993). The threshold is therefore assumed to increase over time, and thus is set as an increasing function of some time parameter, τ_{DFTI}:

$$\theta = f(\tau_{DFTI}) \cdot \sigma, \tag{3.4}$$

where σ is the standard deviation of the error, as defined in Equation 3.2. The goal gradient parameter, c is then defined based on an approach gradient, a, and an avoidance gradient, b:

$$c = b \cdot (v_{R-} + v_{L-}) + a \cdot (v_{R+} + v_{L+}), \tag{3.5}$$

---

[3]Note that in general a different function may be used, but we follow Busemeyer and Townsend (1993) in using the hyperbolic tangent function as it works well for predictions in Table 8 of their paper.

where the mean valence inputs v_R and v_L are split into a weighted sum of 'positive' attributes that are desirable, v_{R+} and v_{L+} (i.e. the average gain for alternative A_R and A_L respectively) and the weighted sum of 'negative' attributes that are undesirable, v_{R-} and v_{L-} (i.e. the average losses). Finally, the mean input valence is defined based on the approach gradient, avoidance gradient, internal threshold and expected gains and losses for each alternative:

$$\delta = (v_{R+} + v_{L+})(1 - a \cdot \theta) + (v_{R-} + v_{L-})(1 - b \cdot \theta), \qquad (3.6)$$

2.2.2 Estimation of the probability of alternatives in DFT with an internal threshold

Whilst the probabilities with which each alternative is chosen under a DFT-I model can be calculated using Markov chain methods (Busemeyer and Diederich, 2002), we instead choose to solve the original integral given by Busemeyer and Townsend (1993), as this is simple to use for the choice scenarios described in this paper[4]. Busemeyer and Townsend (1993) demonstrate that by assuming a continuous time process ($h \to 0$), the probability of choosing alternative A_R over A_L is:

$$Pr(A_R, A_L) = \frac{S(z)}{S(\theta)}, \qquad (3.7)$$

where the function $S(x)$ is the integral:

$$S(x) = \int_{-\theta}^{x} exp\left[\frac{(c+s) \cdot y^2 - 2 \cdot \delta \cdot y}{\sigma^2}\right] dy, \qquad (3.8)$$

with the initial preference value, $P_0 = z$, θ the internal threshold, c the goal-gradient parameter, s the growth-decay parameter, δ the mean input valence and σ^2, the variance of this input, all as defined and specified in Section 2.2.1. To solve this integral, we use the substitution:

$$u = \frac{(c+s) \cdot y - \delta}{\sigma \cdot \sqrt{(c+s)}}. \qquad (3.9)$$

In order to simplify the integral, we square both sides of Equation 3.9 and rearrange to get:

$$\frac{(c+s) \cdot y^2 - 2 \cdot \delta \cdot y}{\sigma^2} = u^2 - \frac{\delta^2}{(c+s) \cdot \sigma^2}. \qquad (3.10)$$

This, together with the derivative:

---

[4]Note that Markov chain methods are more practical to use for more complex scenarios (Bhattacharya and Waymire, 1990; Busemeyer and Townsend, 1993).

$$\frac{du}{dy} = \frac{\sqrt{c+s}}{\sigma},$$ (3.11)

can be used to rearrange the integral in Equation 3.8 in terms of u and du:

$$S(x) = \int_{ll}^{ul} exp\left[u^2 - \frac{\delta^2}{(c+s)\cdot\sigma^2}\right] \cdot \frac{\sigma}{\sqrt{c+s}}\, du$$ (3.12a)

$$= exp\left[-\frac{\delta^2}{(c+s)\cdot\sigma^2}\right] \cdot \frac{\sigma}{\sqrt{c+s}} \cdot \int_{ll}^{ul} exp[u^2]\, du,$$ (3.12b)

where the limits of integration are $ll = \frac{-\theta\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}$ and $ul = \frac{x\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}$. Finally, we note that we require the imaginary error function ($erfi(x)$, Abramowitz and Stegun 1965), which is defined:

$$erfi(x) = \frac{2}{\sqrt{\pi}} \cdot \int_0^x exp[u^2]\, du.$$ (3.13)

Following some rearrangement, this results in a solution to Equation 3.7:

$$Pr(A_R, A_L) = \frac{S(z)}{S(\theta)}$$ (3.14a)

$$= \frac{erfi\left(\frac{z\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}\right) - erfi\left(\frac{-\theta\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}\right)}{erfi\left(\frac{\theta\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}\right) - erfi\left(\frac{-\theta\cdot(c+s)-\delta}{\sigma\cdot\sqrt{(c+s)}}\right)}.$$ (3.14b)

2.3 The impact of time in DFT

Under a decision field theory model, the preferences of alternatives change as the decision-maker deliberates on the attributes of the alternatives. Consequently, the probability of each alternative being chosen changes depending on how long the decision-maker takes to make their choice. Under MDFT, the probabilities of each alternative being chosen are impacted by the number of preference accumulation steps, τ, (which we henceforth refer to as τ_{MDFT}), whilst under DFT with an evidence threshold, the probabilities depend on the threshold, θ, which in turn depends on the time parameter, τ_{DFTI}. For a typical simple setting in transport choice modelling, a decision-maker might need to complete choice tasks where there are two route alternatives, one which is cheaper and slower and another which is more expensive and faster. Figure 3.1 demonstrates how the probabilities that each alternative is chosen in such a choice (where the second alternative is 1 minute faster but 1 Swedish Krona more expensive) might change with an increase in decision

time. Behaviourally, the assumption here is that a longer response time results in more comparisons of the alternatives, which results in more time for the preferred attributes to have an impact on the preferences.
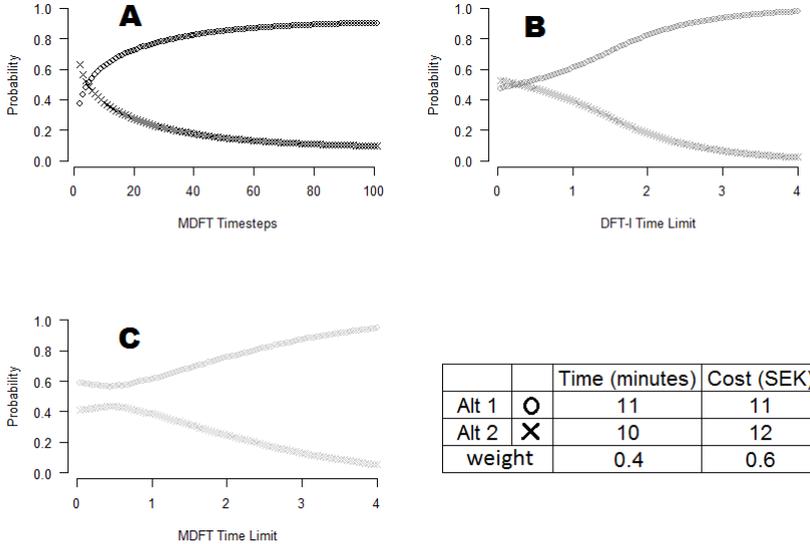
Fig. 3.1: The probability of choosing the two different alternatives as the amount of time taken considering the choices increases.

Graph A shows a choice under MDFT and how the probability of choosing each alternative changes as the number of preference accumulation steps comparing the alternatives increases. Graph B shows the equivalent choice scenario under a DFT-I model, with the probability of alternatives changing with the time parameter (which impacts the value of the internal threshold depending on Equation 3.4). Graph C shows a MDFT model with an evidence threshold. This is created by using an MDFT model as described by Equations 2.1 - 2.8, but rather than calculating expected values and covariances, we run 10,000 simulations of the preferences evolving over time, concluding the choice when the difference in the preference values reaches some 'internal threshold', where the threshold is set as an increasing function of response time (as done for DFT-I). Under each version of DFT, the probability of choosing alternative 1 increases as the number of iterations of preference updating increase.

For MDFT, this figure clearly demonstrates that the higher the number of preference accumulation steps, the more deterministic the decision is[5].

---

[5]Whilst this may not be the case under certain parameter values of MDFT and very specific attribute values, we choose typical attribute values and parameter estimates under MDFT in this scenario.

Given additional comparisons, the fact that the decision-maker is more likely
to consider the cost (which has a weight of 0.6 compared to that of travel
time, which is 0.4), increases the probability of choosing alternative 1, which
is cheaper. Additionally, DFT-I also becomes more deterministic, unsurpris-
ingly, given that 'choice probability becomes more extreme as the threshold
criterion increases' (Busemeyer and Townsend, 1993). Consequently, increas-
ing the time parameter also increases the probability of alternative 1.

There are visible differences in the models with DFT-I producing S-shaped
curves and MDFT producing a more logarithmic-shaped curve. However, re-
sults from simulations of MDFT with an internal threshold produces proba-
bilities more similar to DFT-I, suggesting that the difference between MDFT
and DFT-I is more likely due to the difference in when the choice is made
(time or evidence threshold) rather than the difference in model structure (in
terms of the different parameters the models have). Crucially, independently
of the variation of DFT model used, an increased response time results in
more time for evidence to accumulate, and consequently a more determin-
istic choice. Initially, the impact of the attribute weights is minimal, such
that the chosen alternative depends on the initial preference matrix (which
in Figure 3.1 is in favour of alternative 2) and which attribute is considered
first[6]. As the number of preference accumulation steps increase, the higher
weight for cost begins to have an impact, with the cheaper alternative gradu-
ally becoming more likely to be chosen. For MDFT, we set $\phi_2 = 0.2$ (for the
purpose of this illustration), resulting in the preference values stabilising for
a large number of preference accumulation steps (Berkowitsch et al., 2014),
and hence the probabilities with which each alternative is chosen stabilise
also.

We now look into incorporating choice response time into the models. In
general, the functional form for the number of preference accumulation steps,
$\tau_{MDFT_{np}}$, for individual n in a particular choice task p in a MDFT model
could include multiple parameters:

$$\tau_{MDFT_{np}} = f(R_{np}) + g(z_n) + \epsilon_{np} \tag{3.15}$$

where f is a function of the choice response time, R_{np}, in a particular choice
task, $g(z_n)$ is a function of the characteristics of an individual, n, and ϵ_{np}
is an error term, to be estimated[7]. For example, $g(z_n)$ could capture the
fact that some individuals may not process information as quickly as others,

---

[6]Note that for the applications in this paper, decision-makers are presented with new
alternatives in each trial, therefore there is little argument for the existence of initial
preferences.

[7]In all of the datasets used in this paper, a single choice response time is observed
for each choice task. Thus the deliberation and action components in making the choice
cannot be disentangled. Thus this error term will incorporate 'reaction times' for the
decision-makers to physically select their chosen alternative.

hence they may require a larger number of comparisons before coming to the same conclusions. Whilst previous applications of decision field theory have simply estimated or fixed the number of preference accumulation steps, this paper considers choice tasks where choice response time is recorded. We subsequently consider three distinct methods for incorporating response time into the parameter for the number of preference accumulation steps for MDFT and the time parameter of DFT-I.

For our first set of models (T1), we wish to ensure that the number of preference accumulation steps is positive and that there is at least one step. Thus we define the number of preference accumulation steps as:

$$\tau_{MDFT_{np}} = 1 + e^{(t_0 + t_1 * R_{np})} \tag{3.16}$$

where R_{np} is the choice response time in a particular choice task as before, t_0 is an error term and t_1 is a parameter to be estimated. A positive value for t_1 thus indicates that the number of preference deliberation steps increases with increased response time whereas a negative estimate indicates a decrease in the number of steps for an increased response time[8].

Our second set of models (T2) additionally attempt to utilise differences in response times both across and within individuals. We do this by additionally considering a term t_2 for capturing the impact of a participant's mean response time. T2 therefore has the specification:

$$\tau_{MDFT_{np}} = 1 + e^{(t_0 + t_1 * RSD_{np} + t_2 * log(RM_n))} \tag{3.17}$$

where RM_n is the mean response time for individual n and RSD_{np} is the number of standard deviations the response time for task p is away from RM_n. Again, we use exponentials and add one to ensure that there is at least one step.

Finally, we also try running individual-specific models[9]. Here, for each choice task, we set the number of preference accumulation steps as:

$$\tau_{MDFT_{np}} = 1 + e^{(t_0 + t_1 * RSD_{np})} \tag{3.18}$$

where RSD_{np} is the number of standard deviations away the response time is from the individual's mean response time.

For DFT-I[10], we set the time parameter using the same functions as above but without adding 1 (thus $1 + \tau_{MDFT_{np}} = \tau_{MDFT_{np}}$), as there is no limit

---

[8]Note that in estimation, the number of preference accumulation steps does not have to be a discrete value, as reducing Equation 2.28 allows continuous values to be used for the number of preference updating steps.

[9]Whilst many datasets in large-scale choice modelling applications only have a few observations per individual, one of our datasets has enough such that all observations included in a model can be from just one individual.

[10]Note that we do not run individual models using DFT-I, as the only dataset with enough observations per individual also has three alternatives, and we do not consider simulation of DFT probabilities in the empirical applications of this paper.

on how low the threshold can be (as long as it is positive for one alternative, and negative for the other).

2.4 Latent class models

As it is possible that there exist large differences between individuals, we also use latent class models (Kamakura and Russell, 1989) in this paper (as only one of our three datasets contains enough observations per individual to run individual models). These models allow for differences in sensitivities to be captured, with each different class capturing a different set of taste coefficients, with possibly even different models being used in the different classes (Hess et al., 2012). More often, the same model is used in the different classes, with S different copies of the model (where S is the number of classes) with a different set of parameter estimates β_S estimated for each class. Either way, we can then denote $P_{ni^*t}(\beta_s)$ as the probability of the chosen alternative, i^*, by individual n in choice task t under class s, where β_s is the set of parameters for class s. Allowing for S different classes results in the likelihood of the observed set of choices for individual n is:

$$L_n(\beta, \pi) = \sum_{s=1}^{S} \pi_{ns} \left(\prod_{t=1}^{T_n} P_{ni^*t}(\beta_S) \right) \tag{3.19}$$

where π_{ns} is the estimated share given to model s for participant n (summing to 1 for each n), and T_n is the set of choice tasks faced by the individual.

2.5 Multinomial Logit models

In the empirical applications of this paper, we also wish to compare DFT to static choice models (models in which preferences for alternatives do not change over time for a single set of alternative attributes), for which we use the multinomial logit (MNL) model (McFadden, 1974). This allows us to test the benefits generated by using a model that additionally attempts to capture the choice deliberation process.

Despite the fact that MNL models are static, there are various methods in which response time could be incorporated. Under the assumption of a decision-maker accumulating evidence in favour of alternative, a longer response time would result in a more deterministic choice. As the scale parameter within MNL (which is often normalised to 1) directly controls how deterministic a choice is, a scale parameter set as a function of response time would allow for the possibility of longer response times resulting in more deterministic choices. To illustrate how this would work, we first define our multinomial logit models. If we assume that an individual, n, has a utility, U^*, for alternative j, then:

$$U_{nj}^* = (\beta^*)' x_{nj} + \epsilon_{nj}^* \tag{3.20}$$

where $(\beta^*)'$ is a set of parameters, x_{nj} is vector of observed variables relating to alternative j and ϵ_{nj}^* is the unobserved portion of utility. By assuming a type I extreme value distribution with variance $\sigma^2 \times (\pi^2/6)$, (McFadden, 1974) demonstrates that we can calculate the probability of alternative i being chosen as:

$$P_{ni} = \frac{e^{(\beta^*/\sigma)'x_{ni}}}{\sum_j e^{(\beta^*/\sigma)'x_{nj}}}, \tag{3.21}$$

with σ the scale parameter. Whilst this scale parameter cannot be identified alone from choice data, we can incorporate choice response time into our MNL models using the scale parameter, setting it as we did for the parameter for the number of preference accumulation steps in MDFT and the time parameter in DFT-I. Whilst the scale parameter does not need to be greater than one, it does need to be positive. Additionally we want the impact of t_1 and t_2 to have the same effect on MNL compared to DFT (higher estimates resulting in a more deterministic choice), so we set:

$$\sigma_{np} = \frac{1}{e^{(t_0+t_1*RSD_{np}+t_2*log(RM_n))}}, \tag{3.22}$$

with RSD_{np} and RM_n defined as before.

3 Empirical applications

We will now demonstrate how the response time taken for choosing an alternative can be used in models applied to three very different datasets. All three datasets come from stated choice surveys, where participants consider several sets of alternatives amongst which they have to state their preference. Each choice task comprises of a set of hypothetical alternatives with differing attribute levels. All three questionnaires were completed on a computer, thus response times were recorded automatically when the participant selected an alternative.

3.1 Datasets

3.1.1 Route choice

The route choice dataset tested in this paper comes from a study on choice response time patterns in an online stated choice experiment (Börjesson and Fosgerau, 2015). In each choice task, respondents have two alternative car routes described by travel time and travel cost (in Swedish Krona). We discard choices with a recorded response time of 0 seconds and those with a response time of more than 60 seconds, as we assume that the respondent was either not attempting to respond to the choice seriously or was interrupted. Additionally, we omit choices made by respondents who have less than six (out of eight) choice tasks remaining after the above censor. This leaves us with 15,546 choice tasks completed by 2,358 respondents.

3.1.2 Conservation choice

The conservation dataset tested in this paper comes from a study exploring tree planting preferences in a stated choice survey (Mahieu et al., 2016). 146 participants completed 16 stated choice tasks where they were asked which of two conservation programmes they preferred. Each of the programmes are described by four attributes: country (Senegal or Peru), provision of online information (Yes/No), type of programme (restorative or preservative) and cost (2,5,10,15 EUR). Country and type of programme are both found to have an insignificant impact on the choice and are therefore omitted in this study. In all tasks, the participant also have a third 'status quo' alternative where they could choose not to invest in either of the presented programmes. We again exclude choices with a recorded response time of less than one second or choices with a time of more than a minute. This leaves us with 2,334 (out of 2,336) choice tasks.

3.1.3   Accommodation choice

Our accommodation choice dataset has a total of 32 participants each completing 45 choice tasks. It comes from Cohen et al. (2017)'s paper in which a version of MDFT is fitted with a fixed number of preference accumulation steps, $\tau_{MDFT} = 500$. In each choice task, decision-makers had three accommodation alternatives described by ease of transportation, size, condition and kitchen facilities (each on a scale of 1-5). Figure 3.2 gives an illustration of an example choice scenario. Trials were omitted if the decision-maker did not consider more than two out of twelve of the information panels as Cohen et al. (2017) deemed these to not be meaningful decisions[11]. Consequently, this leaves a total of 1,430 decisions.

	Ease of Transportation	Size	Condition	Kitchen Facilities
Apartment 1	☆☆☆★★	☆☆★★★	☆☆★★★	☆☆☆☆☆
Apartment 2	☆★★★★	☆☆☆☆☆	☆☆★★★	☆☆★★★
Apartment 3	☆☆☆★★	☆☆★★★	☆☆☆☆★	☆☆☆★★

Fig. 3.2: An example choice scenario from Cohen et al. (2017)

3.2   Basic models

We initially do not consider choice response time, simply using basic MDFT and DFT-I models to test the importance of the parameter for the number of preference accumulation steps and time parameter respectively. Given that the parameter for the number of preference accumulation steps is often fixed to a high value, these models effectively test how much flexibility is gained by freeing this parameter. We run five versions of MDFT on each dataset and five versions of DFT-I on just the route choice dataset, as this is the only dataset with only two alternatives.

In the first four versions, we fix the number of preference accumulation steps to 1, 10, 100 and 1,000, respectively for MDFT. For DFT-I, we fix the time parameter to 1, 2, 3, and 4. These values have no link with 'real time' in that the response time for each choice is not included, thus the different values simply test the parameter impact on the model. In the final

---

[11]Note that this was possible as the original experiment used eye-tracking equipment. See Cohen et al. (2017) for details.

version of each model for each dataset, we simply estimate the number of steps/time parameter to be some constant, $\tau_{MDFT} = \tau_{DFTI} = t_0$ (we are not yet including choice response time).

To estimate the probability of alternatives under a decision field theory model with a time threshold, we require estimates for $n - 1$ weight parameters (where n is the number of attributes) and estimates for four process parameters (ϕ_1 and ϕ_2, the sensitivity and memory parameters respectively, the constant for the number of preference accumulation steps, t_0, and the standard deviation of the error term, σ_ϵ).

We additionally consider new MDFT parameters equivalent to alternative specific constants in MNL (see Table 2.4 in Chapter 2). These parameters are used to capture underlying preferences for alternatives and are particularly useful if there is a status-quo bias or an unfavourable alternative (such as not picking either conservation programme in our conservation choice dataset). For the work in this paper, we use additional attribute weights for time spent considering a specific alternative, j. For example, dummy attributes of $(0, 0, 1)$ could be used to represent 'additional time spent considering alternative 3 not captured by the attribute values'. These additional dummy attributes are omitted for the housing dataset, for which they are not found to be significant. We choose to use additional attribute weights rather than parameters in the initial preference matrix as we wish to test the importance and relative impact of the number of preference accumulation steps, which would be affected by an initial preference matrix.

For DFT-I, which is tested only on the route choice dataset, we also estimate two attribute weight parameters, with a third dummy attribute of $(0, -1)$ used to represent additional time spent considering the negatives of alternative 2. This is equivalent to $(1, 0)$ but we use a negative as this results in all three attributes being negative, meaning that we do not need to estimate an approach gradient, a, and thus only need to estimate b, the avoidance gradient. However, we set $b = 0$ and the growth-decay rate $s = 0$ as this does not significantly impact model fit. Additionally, we set the bias term $z = 0$ as, similarly to MDFT, we wish to avoid impacting the time parameter. This results in only two free parameters (the attribute weights) for DFT-I models with a fixed time parameter.

For all models, the free parameters are estimated by maximisation of the likelihood function of the observed choices. We use the R packages maxLik (Henningsen and Toomet, 2011) and cmcRcode (CMC, 2017) for estimation of the likelihood function. We use an RCPP package together with the Armadillo C++ linear algebra library for calculation of the matrices required for finding the probability of alternatives in the MDFT models (Eddelbuettel et al., 2011; Sanderson and Curtin, 2016). Finally, we use a initial parameter search algorithm based on Bierlaire et al. (2010)'s heuristic for non-linear global optimisation to try minimise the risk of our model finding poor local

optima. The results of the basic DFT models are shown in Table 3.1.

Table 3.1: The number of free parameters (f.p), log-likelihoods (LL) and estimate/fixed value for the time parameter (τ_{DFTI}) or number of preference accumulation steps (τ_{MDFT}) for the basic DFT models

| | | DFT Evidence Threshold | | | DFT Time Threshold | |
| | | Route | | | Route | |
	f.p.	LL	τ_{DFTI}	f.p.	LL	τ_{MDFT}
restricted	2	-7,256.77	*1*	5	-7,441.19	*1*
restricted	2	-6,932.33	*2*	5	-6,904.70	*10*
restricted	2	-7,203.34	*3*	5	-6,915.92	*100*
restricted	2	-7,827.38	*4*	5	-6,915.92	*1,000*
τ free	3	**-6,930.75**	1.92	6	**-6,883.18**	5.83
		DFT Time Threshold			DFT Time Threshold	
		Accommodation			Conservation	
	f.p.	LL	τ_{MDFT}	f.p	LL	τ_{MDFT}
restricted	6	-1,429.20	*1*	6	-2,097.82	*1*
restricted	6	-1,329.75	*10*	6	-1,959.26	*10*
restricted	6	-1,320.69	*100*	6	-1,961.03	*100*
restricted	6	-1,330.40	*1,000*	6	-1,961.03	*1,000*
τ free	7	**-1,320.59**	290.54	7	**-1,959.17**	8.96

For all three datasets, DFT achieves a better model fit if the parameter for the number of preference accumulation steps/time parameter is free. It is notable that MDFT models for both route and accommodation choice with a free parameter for the number of preference accumulation steps have vastly better fit than models with the number of steps fixed to 1,000, as often previously configured. Additionally, the estimate for the number of preference accumulation steps varies considerably across the datasets, further demonstrating the value of having a freely estimated parameter for the number of preference accumulation steps. DFT-I similarly produces results that are vastly worse if the time parameter is inappropriately fixed.

3.3   Models with choice response time

We next set the number of preference accumulation steps, τ_{MDFT} and the time paramater, τ_{DFTI} as functions of choice response time, using Equations 3.16 and 3.17 for models T1 and T2, respectively. For all three datasets, there is a statistically significant improvement in model fit by incorporating choice response time (see Table 3.2).

Whilst incorporating response time alone has little impact except in our conservation dataset (cf. models T1), where in general a faster decision is less deterministic, the real gain of using response time in our MDFT models is seen in models T2. This results in a slower mean response time indicating a more deterministic response ($t_2 > 0$, see Table 3.2), a finding that is consis-

Table 3.2: Results from including choice response time in DFT, with log-likelihoods and estimates and robust t-ratios for our time parameters.

Dataset/Model			T0	T1	T2
		Log-likelihood	-6,930.75	-6,927.20	-6,924.73
	t_0	estimate	0.65	0.60	0.27
		rob. t-ratio	24.81	18.71	1.56
Route (Evidence Threshold)	t_1	estimate	-	3.4E-03	0.02
		rob. t-ratio	-	2.82	1.99
	t_2	estimate	-	-	0.15
		rob. t-ratio	-	-	2.31
		Log-likelihood	-6,883.18	-6,882.37	-6,874.37
	t_0	estimate	1.58	1.50	0.16
		rob. t-ratio	12.04	9.85	0.36
Route (Time Threshold)	t_1	estimate	-	0.01	-0.02
		rob. t-ratio	-	1.35	-0.58
	t_2	estimate	-	-	0.54
		rob. t-ratio	-	-	3.22
		Log-likelihood	-1,320.59	-1,320.38	-1,307.36
	t_0	estimate	5.67	6.25	4.40
		rob. t-ratio	2.55	4.40	4.85
Accommodation (Time Threshold)	t_1	estimate	-	-0.01	-0.44
		rob. t-ratio	-	-0.45	-2.75
	t_2	estimate	-	-	0.72
		rob. t-ratio	-	-	3.53
		Log-likelihood	-1,959.20	-1,919.20	-1,917.96
	t_0	estimate	2.07	-4.73	-11.18
		rob. t-ratio	6.53	-1.50	-1.82
Conservation (Time Threshold)	t_1	estimate	-	2.08	1.05
		rob. t-ratio	-	2.42	1.62
	t_2	estimate	-	-	8.51
		rob. t-ratio	-	-	1.98

tent across all three datasets (for both MDFT and DFT-I). Additionally, an individual making a decision faster than their average response time is more deterministic ($t_1 < 0$) for that choice in some cases and less deterministic ($t_1 > 0$) in others. Consequently, by including both the decision maker's response time in comparison to others and to themselves, MDFT model fit is consistently improved across all three datasets. However, DFT-I has positive significant estimates for both t_1 and t_2, suggesting that a longer response time relative to either an individual's mean response time (cf. t_1) or a longer mean response time (cf. t_2) result in more deterministic choices.

These results are visualised through Figure 3.3, which shows the estimates for the number of preference accumulation steps and the time parameter for each choice, depending on the decision-maker and their response times for each individual choice.
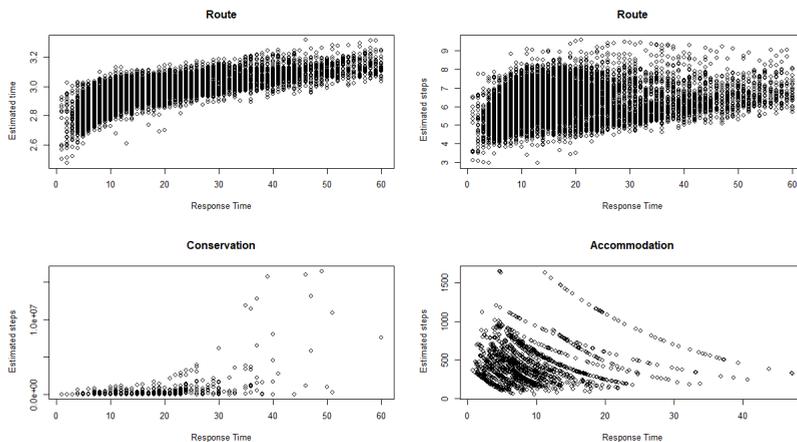
Fig. 3.3: The estimates for the number of preference accumulation steps or the time parameter depending on the response times

Whilst the estimated number of preference accumulation steps is extremely high (and implausible from a psychological perspective) for the conservation dataset, it is worth noting that the estimate for the memory parameter is $\phi_2 = 0.043$, meaning that the probabilities with which each alternative is chosen converges after a large number of preference accumulation steps. Consequently, if all decisions where $\tau_{MDFT} > 1000$ instead have 1000 steps, there is no change in model fit. This does however suggest that there is a large difference in how deterministic the different decision-makers are, as small changes for very fast responses have a significant impact.

3.4 Individual response time models

As we have up to 45 choice tasks per individual in our accommodation choice dataset, we can also run a MDFT model for each individual. This allows us to directly explore whether including response time aids model fit for explaining the choices made by each individual decision-maker. The results of these models are displayed in Figure 3.4. In line with the overall model (T2 in Table 3.2), we see that most estimates for t_1 are negative, resulting in decision-makers typically being more deterministic if they make a faster response. Whilst in many cases the estimate for t_1 is insignificant, in 10 out of 32 cases the log-likelihood for the model fit improves by more than 5% (with 8 of these 10 having a negative estimate for t_1 and the other 2 having a positive estimate). In 4 cases (all of which have negative estimates for t_1) the improvement is between 10 and 15%.
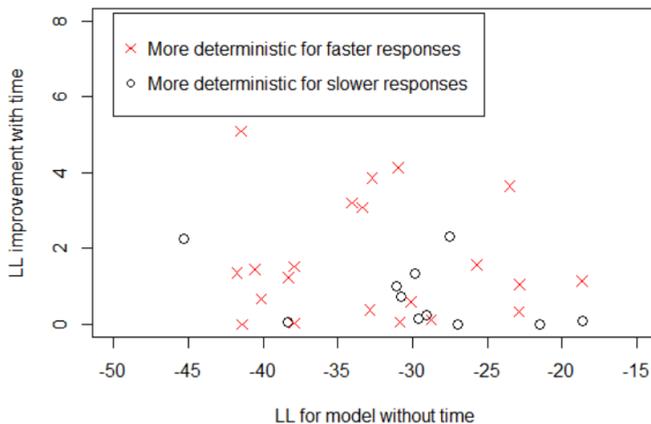
Fig. 3.4: The log-likelihoods for individual MDFT models compared to the improvement in fit by including response times

3.5 Latent class models

Given that some individuals might be more deterministic for faster responses and some might be more deterministic for slower responses, we also try latent class models to test whether separating the participants into two classes helps improve our models. Whilst we could use a function of parameters such that the class allocation is dependent on characteristics of an individual, in this case we are interested in whether there are different types of respondents based on their response time. Thus, in our first set of latent class models, we do not use response time and simply test the impact of an MDFT model with two different values for the number of preference accumulation steps and a DFT model with two different values for the time parameter (denoted τ_1 and τ_2 in Table 3.3 below). This means that the models only have two additional parameters, one for a second estimate for the number of preference accumulation steps, and one for the class share. The rest of the parameter estimates, including the attribute weights, are held constant across the two classes. The results are shown in Table 3.3.

Table 3.3: Results for basic latent class DFT models for the three datasets

		classes	f.p.	LL	BIC	τ_1	class 1 share	τ_2
DFT-I		1	3	-6930.75	13,890.45	1.92	100%	
		2	5	-6745.06	13,538.37	6.37E-04	9%	2.23
	Route	1	6	-6,883.18	13,824.27	5.83	100%	
		2	8	-6,742.23	13,561.67	1.00	20%	10.63
MDFT	Accommodation	1	7	-1,320.59	2,692.03	290.54	100%	
		2	9	-1,310.15	2,685.68	789.00	35%	4,704.68
	Conservation	1	7	-1,959.20	3,972.68	8.96	100%	
		2	9	-1,830.76	3,731.31	1.00	31%	19.28

76

Notably, the use of latent classes results in better fit for all three datasets, with a lower Bayesian Information Criteria (BIC) value for all datasets. For the route and conservation datasets, it appears that a subset of the decision-makers (9% for DFT-I or 20% for MDFT and 31% for MDFT, respectively for route and conservation) are much more random in their choices, with an estimate for the number of preference accumulation steps of 1 and a time parameter of 0 being found. This vastly improves model fit for both these datasets. This suggests that these individuals have very different sensitivities to the attributes in comparison to the majority of individuals, as the estimates suggest that random choices (where there is little/no time for evidence to accumulate) better represent their choices[12].

Given these results, it appears that our latent class models should also incorporate a separate set of attribute weights for each class (version 2 in Table 3.4). We can additionally also look at the impact of having two different classes for the DFT models whilst also including the response times taken by the decision-makers to make their choices. For these models (version 3 in Table 3.4), we use Equation 3.17 to set the number of preference accumulation steps/the time parameter, with different values for t_0, t_1 and t_2 in the two different classes. The results of latent class models 2 and 3 are given in Table 3.4.

Crucially, there is a vast improvement for all models in comparison to the first set of latent class models, demonstrating that there is a large amount of taste heterogeneity across all three datasets going beyond just the τ_1/τ_2 split. For the route choice dataset, it appears that some individuals are more sensitive to time (wt_1), whilst others are more sensitive to cost (wt_2). The key differences are the cost for the accommodation (wt_2) and the bias against picking neither alternative (wt_4) for conservation datasets respectively. Whilst the log-likelihoods for latent class models with response time are similar to those of latent class models without response time, there is still a significant improvement for models across all three datasets (though not in terms of BIC for route and conservation). The parameter estimates for the response time coefficients in both classes are shown in Table 3.4 also, with many positive and negative coefficients. Figure 3.5 shows this effect clearly, with, for example, the time parameter increasing with response time for class 1 for DFT-I, but the time parameter decreasing with response time for class 2. For the conservation dataset it appears that there is one group (class 2) where response time does not influence randomness, with the other (class 1) producing less deterministic choices with increasing response time. Both classes produce less deterministic choices with increasing response time for the accommodation dataset.

---

[12]Note that the wide range in the number of preference accumulation steps for the model incorporating response time for the conservation dataset also implies this.

Table 3.4: The results of versions 2 and 3 of latent class models for both variations of DFT

Model	DFT Evidence Threshold				DFT Time Threshold			
Dataset	Route				Accommodation		Conservation	
Version	2	3	2	3	2	3	2	3
response time	no	yes	no	yes	no	yes	no	yes
free parameters	7	11	13	17	15	19	15	19
Log-likelihood	-6,017.46	-6,007.09	-5,983.38	-5,976.97	-1,255.22	-1,238.64	-1,745.76	-1,734.77
BIC	-12,102.49	-12,120.34	-12,092.22	-12,118.02	-2,619.42	-2,615.31	-3,521.53	-3,616.89
Class 1 Par. Ests.								
class 1 share	0.62	0.62	0.64	0.63	0.59	0.69	0.75	0.73
t_0	0.82	1.07	1.94	2.45	7.71	5.66	2.43	1.87
t_1	-	-0.02	-	0.03	-	-0.34	-	-0.38
t_2	-	-0.02	-	-0.19	-	-0.93	-	-0.36
wt_1	0.25	0.25	0.26	0.25	0.11	0.13	0.04	0.03
wt_2	0.62	0.63	0.64	0.65	0.47	0.42	0.10	0.08
wt_3	0.13	0.12	0.10	0.10	0.14	0.17	0.11	0.13
wt_4	-	-	-	-	0.28	0.29	0.75	0.76
Class 2 Par. Ests.								
class 2 share	0.38	0.38	0.36	0.37	0.41	0.31	0.25	0.27
t_0	1.10	0.66	7.28	-2.43	5.64	6.90	1.03	-1.21
t_1	-	3.00E-04	-	-0.02	-	-0.82	-	-0.01
t_2	-	0.03	-	2.11	-	-0.20	-	-0.93
wt_1	0.51	0.50	0.49	0.48	0.20	0.20	0.00	0.00
wt_2	0.29	0.28	0.26	0.26	0.38	0.33	0.22	0.20
wt_3	0.20	0.22	0.25	0.26	0.30	0.35	0.39	0.37
wt_4	-	-	-	-	0.12	0.12	0.39	0.43
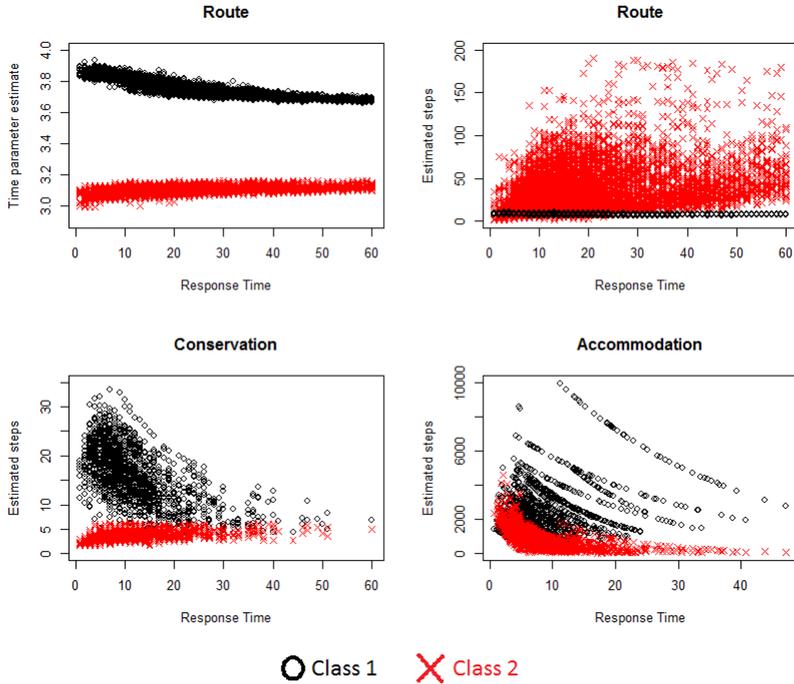
Fig. 3.5: The estimates for the number of preference accumulation steps and the time parameter depending on the response times for the latent class models.

3.6 Reappraisal of results using multinomial logit models

As our results demonstrate that DFT time parameters can be successfully parameterised as a function of response time, this leads us to consider whether alternative models can similarly include response time information. The most obvious example of a model to test is the multinomial logit (MNL) model (McFadden, 1974). Whilst this is a static model, the response time can be incorporated easily through the use of Equation 3.22.

We now test five different MNL models for each dataset, with each one being equivalent to a DFT models tested already in this paper.

1. A basic MNL model without response time ($\sigma = 1$ in Equation 3.22).

2. A MNL model with response time captured as defined in Equation 3.22.

3. A MNL model with 2 latent classes, with only the scale parameter different and without response time

79

4. A MNL model with 2 latent classes, with different scale and weight parameters, but without response time

5. A MNL model with 2 latent classes, each with a complete set of separate parameters, including separate response time parameters t_1 and t_2.

The results from our MNL models are shown in Table 3.5.

Whilst MNL has vastly worse fit for the route choice dataset, the MNL models have very similar log-likelihoods to their corresponding DFT models[13] for the conservation and accommodation choice datasets. In particular, the impact of including response time is similar for MNL compared to DFT.

Consequently, it appears that adjustments in the specification for the scale parameter in MNL have very similar impacts to the equivalent adjustments for the time parameters in DFT. Additionally, the time parameter estimates for the single class version of MNL and MDFT for all three choice sets appear to be similar in that they always have the same sign (see Table 3.6). The t_1 and t_2 estimates for MNL are approximately half that of those for the corresponding MDFT models. The impacts of this are particularly clear when considering Figure 3.6, in which the distributions for the scale parameter estimates for the route and accommodation datasets very closely resemble that of the estimated number of preference accumulation steps for MDFT in Figure 3.3.
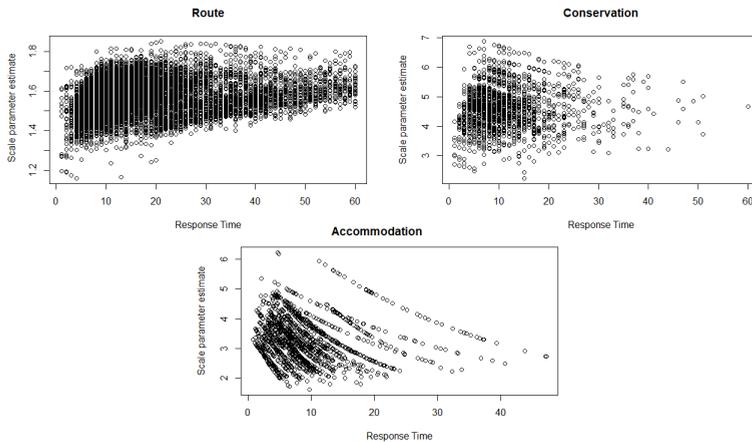
Fig. 3.6: The scale parameter estimates in the MNL models depending on the response times.

This implies that the parameter for the number of preference accumulation steps in a MDFT model can perform the same function as the scale

---

[13]Note that DFT-v2 is the same model as DFT-T2 in Table 3.2.

Table 3.5: Results from the MNL models, with log-likelihoods also shown for equivalent DFT models

model	version	response time	weights difference	classes	Route f.p	Route LL	Conservation f.p	Conservation LL	Accommodation f.p	Accommodation LL
MNL	1	no	no	1	3	-8,769.75	4	-1,958.68	4	-1,329.09
	2	yes	no	1	5	-8,767.11	6	-1,934.59	6	-1,313.95
	3	no	no	2	5	-7,395.42	6	-1,841.38	6	-1,316.49
	4	no	yes	2	7	-7,105.15	9	-1,749.63	9	-1,265.66
	5	yes	yes	2	12	-7,089.40	14	-1,739.94	14	-1,257.27
DFT Time Threshold	1	no	no	1	6	-6,883.18	7	-1,959.20	7	-1,320.59
	2	yes	no	1	8	-6,874.37	9	-1,917.96	9	-1,307.36
	3	no	no	2	8	-6,742.23	9	-1,830.76	9	-1,310.15
	4	no	yes	2	13	-5,983.38	15	-1,745.76	15	-1,255.22
	5	yes	yes	2	17	-5,976.97	19	-1,734.77	19	-1,238.64
DFT Evidence Threshold	1	no	no	1	3	-6,930.75	-	-	-	-
	2	yes	no	1	5	-6,924.73	-	-	-	-
	3	no	no	2	5	-6,745.06	-	-	-	-
	4	no	yes	2	7	-5,986.41	-	-	-	-
	5	yes	yes	2	11	-5,976.97	-	-	-	-

parameter in a MNL model. The consequences of this are that if we only have the choices and response times of a decision-maker, then MDFT may not make any further gain than a far simpler MNL model (with the exception being the single class version for the conservation dataset, although this gain is lost once two classes are considered). This suggests that attempting to capture the deliberation process, in this case, does not add anything to the performance of the model.

Whilst the same cannot be said of MNL and DFT-I, which have somewhat different estimates (see Table 3.6), the results for DFT-I suggest that response time cannot be directly linked the time parameter. This is because the latent class DFT-I model has a significant negative estimate for t_2 for one of the classes, suggesting that individuals in this class are more deterministic if they have a faster mean response time.

4   Conclusions

The work in this paper was motivated by the recent improvements (in Chapter 2) in the computational mechanisms underlying decision field theory models with an external threshold (MDFT). With it now being easily possible to incorporate decision response time into a MDFT model, this paper considers the impact of this on three datasets, as well as testing the inclusion of response time in models based on the original specification of DFT (labelled DFT-I in this paper) by Busemeyer and Townsend (1993), in which individuals conclude a decision when they reach an internal (evidence) threshold.

In favour of the notion of preference accumulation, we find that MDFT models estimate a larger number of preference accumulation steps for decision-makers who have a longer mean response time relative to other decision-makers. However, contradicting the notion of evidence accumulation, we find some negative estimates for the impact of increased response time meaning that more deterministic decisions are made in choices where a decision-maker responds more quickly than their mean response time across all of their choice tasks. Additionally, the impact of the average response time for an individual diminishes with latent class MDFT models, with two separate estimates for the number of preference accumulation steps controlling for the effect of a longer mean response time. With MDFT parameters for the number of preference accumulation steps behaving very similarly to MNL scale parameters, it appears that the step parameter cannot be directly linked to response time. This means that a MDFT model with response time may not in fact capture a choice deliberation process. This is perhaps unsurprising given that MDFT models could be considered inappropriate for the choice tasks in this paper, for which no time limit is imposed (contrary to the key assumption of MDFT, that a decision is made upon reaching a specific time limit). However, we do

Table 3.6: Time parameter estimates from DFT and MNL models

Dataset			MNL LC1	MDFT LC1	DFT-I LC1	MNL LC2		MDFT LC2		DFT-I LC2	
Route	Log-likelihood		-8,767.11	-6,874.37	-6,924.73	-7,089.40		-5,976.97		-6,007.09	
	t_0	estimate	0.00	0.16	0.27	0.00	-3.95	2.45	-2.43	1.07	0.66
		rob. t-ratio	fixed	0.36	1.56	fixed	-4.53	4.39	-1.18	10.77	4.45
	t_1	estimate	-0.01	-0.02	0.02	-0.13	-0.06	0.03	-0.02	-0.02	0.00
		rob. t-ratio	-0.41	-0.58	1.99	-2.56	-1.12	0.45	-0.19	-0.74	0.01
	t_2	estimate	0.17	0.54	0.15	0.12	0.56	-0.19	2.11	-0.02	0.03
		rob. t-ratio	2.17	3.22	2.31	0.64	2.13	-0.92	2.43	-2.67	3.18
Accommodation	Log-likelihood		-1,313.95	-1,307.35	-	-1,257.27		-1,238.64		-	
	t_0	estimate	0.00	4.40	-	0.00	0.16	5.66	6.90	-	-
		rob. t-ratio	fixed	4.85	-	fixed	0.07	7.64	1.22	-	-
	t_1	estimate	-0.28	-0.44	-	-0.17	-0.23	-0.34	-0.82	-	-
		rob. t-ratio	-4.58	-2.75	-	-1.68	-1.58	-2.10	-3.30	-	-
	t_2	estimate	0.36	0.72	-	0.04	0.34	0.93	-0.20	-	-
		rob. t-ratio	3.11	3.53	-	0.04	1.02	3.63	-0.25	-	-
Conservation	Log-likelihood		-1,934.59	-1,917.96	-	-1,739.94		-1,734.77		-	
	t_0	estimate	0.00	-11.18	-	0.00	0.38	1.87	-1.21	-	-
		rob. t-ratio	fixed	-1.82	-	fixed	0.33	2.04	-0.48	-	-
	t_1	estimate	-0.08	1.05	-	-0.22	-0.15	-0.38	-0.01	-	-
		rob. t-ratio	-2.09	1.62	-	-0.74	-1.31	-3.09	-0.03	-	-
	t_2	estimate	0.55	8.51	-	-0.05	0.23	0.36	0.93	-	-
		rob. t-ratio	3.06	1.98	-	-0.03	1.16	0.88	0.65	-	-

not computationally impose a strict time limit on our MDFT models (with the number of preference accumulation steps being adjusted according to the response times) and results from our DFT models with internal thresholds are also unfavourable for the notion of preference accumulation. Our latent class DFT-I model results in both positive and negative estimates for the effect of a participant's mean response time on how deterministic the choice is. This implies that the core assumption of DFT, that choice probabilities become more extreme with a higher threshold (evidence or time), cannot be predicted when considering response times. However, the impact of response time on the parameter for the number of preference accumulation steps and the time parameter might look very different if a DFT model could control for choice certainty in a more direct way. Thus, further work could consider additionally linking choice certainty as well as response times to the time parameters.

Whilst our results suggest that the deliberation process may not be truly captured by an DFT model, this work does provide evidence that incorporating response times into choice models can improve model fit. The fact that response time can be directly included in calculating the probability of alternatives may have some impact in stated preference studies but is far more likely to have an impact in work involving revealed preference data, where decisions such as what to order in a restaurant, who to vote for and which route to choose at a roadway junction are more likely to be impacted by time pressure due to the nature of the choices. Additionally, previous work has demonstrated that the linear ballistic accumulator model (Brown and Heathcote, 2008), can be adjusted such that changing information can be incorporated into the model (Holmes et al., 2016). Similarly, it is possible that as well as incorporating response time, a decision field theory model could incorporate changing attributes, such as in a dynamic price setting (e.g. auctions, flight booking websites) or travel times for different routes. This could also prove useful for studying how commuters change their route choice when forced to do so due to a change to their original schedule, such as a delayed train. In particular, dynamic models such as DFT may prove useful for econometric forecasting, particularly if we do not have much information on the decision-maker but there is some indication on how long they might take to make the decision. However, DFT and other accumulation models may only have advantages over basic models such as multinomial logit by considering such data complexities, without which, results from this paper suggest that MNL can perform just as well for multi-alternative, multi-attribute choice. Thus, a DFT model for dynamic data may also require information such as choice certainty if it is to capture the deliberation process accurately.

Further work could also consider latent constructs, where latent variables are used to predict both response times and the number of preference accumulation steps. Certain individuals may process information at different

rates, meaning that there is likely to be variation in the estimated number of iterations of preference updating per second across individuals. It may not be that individuals who spend longer considering a decision are necessarily considering the alternative in more detail, as implied by the structure we impose on the models in this application. Using random parameters to capture this difference across individuals therefore may have much more explanatory power than when point estimates for parameters are used.

Whilst these results give various implications and provide many directions for future work, it is clear that, where possible, analysts should record response times in choice decisions, as it is likely that both dynamic models such as DFT and simpler models such as multinomial logit are able to use this information to better predict the choices made. This may lead to more accurate estimations and forecasts in many situations.

Acknowledgements

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

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.

Bhattacharya, R. N. and Waymire, E. C. (1990). *Stochastic Processes with Applications*, volume 61. SIAM.

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

Börjesson, M. and Fosgerau, M. (2015). Response time patterns in a stated choice experiment. *Journal of choice modelling*, 14:48–58.

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

References

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.

CMC (2017). CMC choice modelling code for R, Choice Modelling Centre, University of Leeds, UK.

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. and Zachary, S. (1978). Improved multiple choice models. *Determinants of travel choice*, 335:357.

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.

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

Hess, S., Stathopoulos, A., and Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591.

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.

Kamakura, W. A. and Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of marketing research*, pages 379–390.

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

Mahieu, P.-A., Crastes, R., Louviere, J., Zawojska, E., et al. (2016). Rewarding truthful-telling in stated preference studies. Technical report.

References

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

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

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.

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.

Raab, M. and Johnson, J. G. (2004). Individual differences of action orientation for risk taking in sports. *Research quarterly for exercise and sport*, 75(3):326–336.

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. and Black, I. R. (2006). Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters*, 17(4):295–310.

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

Schall, J. D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Current opinion in neurobiology*, 13(2):182–186.

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.

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.

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, 117(4):1275.

Turner, B. M., Schley, D. R., Muller, C., and Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, 125(3):329.

References

Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.

Chapter 4

An accumulation of preference: contrasts between Decision Field Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for travel behaviour modelling

Thomas Hancock[1], Stephane Hess[1], & Charisma Choudhury[1]

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-*

---

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)

Chapter 4. An accumulation of preference: contrasts between Decision Field
Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for
travel behaviour modelling

*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 esti-
mation 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 be-
haviour in other fields has been implemented with very different aims and
objectives. Since work in the 1970s (Tversky, 1972, 1977; Tversky and Kah-
neman, 1973), 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 mod-
els. Choice modellers have long had an interest in increasing the behavioural
realism of their models, with recent methodological advances aimed at incor-
porating alternative behavioural ideas such as random regret minimisation
(RRM) (Chorus, 2010; Chorus et al., 2008), 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 (Noguchi and Stewart, 2014;
Roe et al., 2001; Trueblood et al., 2013b) 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) and work in Chapter 2). 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(in Chapter 2). 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 (Cohen et al., 2017; Trueblood et al., 2014; Turner et al., 2017)[2].

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 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

---

[2]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).

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[3]. 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; Krajbich et al., 2012; Usher and McClelland, 2001). 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.

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

---

[3]For further details and a more comprehensive description of decision field theory, please refer to Section 2 in Chapter 2.

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 in Chapter 2 of this thesis. 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 (González-Valdés and Ortúzar, 2017; Kaufman, 1990; Schwartz et al., 2002)) 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 4.1 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.

Fig. 4.1: An example decision process under both accumulation models
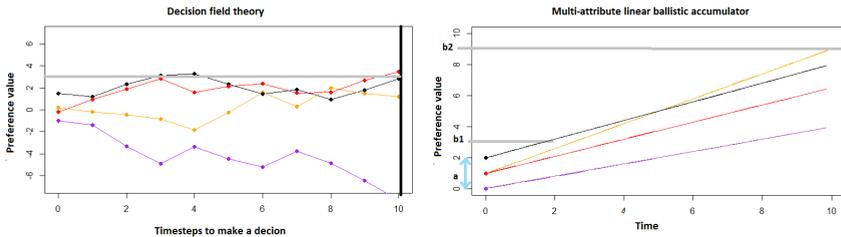
The linear ballistic accumulator model (Brown and Heathcote, 2008) and its multi-attribute version MLBA (Trueblood et al., 2013a, 2014) have a similar accumulation process for the preference of alternatives, but, in contrast

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 4.1, 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 4.1 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

Chapter 4. An accumulation of preference: contrasts between Decision Field
Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for
travel behaviour modelling
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

For a full description of the theory and estimation of DFT models, readers should refer to Section 2 and 3 of Chapter 2.

2.2.2 Multi-attribute linear ballistic accumulator

Theory

Under MLBA, each alternative has a value that linearly grows towards a threshold (see right panel in Figure 4.1). The chosen alternative in an MLBA model is the first alternative to pass a threshold value, χ. 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 4.1 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 B). 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:

$$D_j \sim TN\left(d_j, s_j\right) \tag{4.1}$$

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):

$$d_j = v_j + I_0 \tag{4.2}$$

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:

$$v_j = \sum_{j \neq i} \sum_{k=1}^{K} \left(w_{x_{k,i,j}} \cdot (x_{k,j} - x_{k,i}) \right). \tag{4.3}$$

In this notation, $x_{k,i}$ is the value for the kth 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:

$$w_{x_{k,i,j}} = exp(-\lambda \cdot |x_{k,i} - x_{k,j}|) \tag{4.4}$$

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

$$\lambda = \begin{cases} \lambda_1, & \text{if } x_{k,j} \geq x_{k,i}. \\ \lambda_2, & \text{if } x_{k,j} < x_{k,i}. \end{cases} \tag{4.5}$$

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 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.

[4]Note that Equation 4.3 is equivalent to Equation 3 in Trueblood et al. (2014), but for multiple alternatives and multiple attributes.

Estimation of MLBA

If we have values (either estimated or fixed) for the drift rates of the alternatives and for the start and end points (A and χ 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 χ 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[χ-A,χ] (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:

$$F_j(t) = Prob\left(\frac{U[\chi - A, \chi]}{D_j} < t\right) \tag{4.6}$$

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

$$F_j(t) = 1 + \frac{\chi - A - t \cdot D_j}{A} \cdot \Phi\left(\frac{\chi - A - t \cdot D_j}{t \cdot s}\right) - \frac{\chi - t \cdot D_j}{A} \cdot \Phi\left(\frac{\chi - t \cdot D_j}{t \cdot s}\right)$$
$$+ \frac{t \cdot s}{A} \cdot \phi\left(\frac{\chi - A - t \cdot D_j}{t \cdot s}\right) - \frac{t \cdot s}{A} \cdot \phi\left(\frac{\chi - t \cdot D_j}{t \cdot s}\right) \tag{4.7}$$

where ϕ and Φ are the standardised normal distribution's density and cumulative density functions, respectively. The associated probability density function is then:

$$f_j(t) = \frac{1}{A}\left[- D_j \cdot \Phi\left(\frac{\chi - A - t \cdot D_j}{t \cdot s}\right) + D_j \cdot \Phi\left(\frac{\chi - t \cdot D_j}{t \cdot s}\right)\right.$$
$$\left. + s \cdot \phi\left(\frac{\chi - A - t \cdot D_j}{t \cdot s}\right) - s \cdot \phi\left(\frac{\chi - t \cdot D_j}{t \cdot s}\right)\right] \tag{4.8}$$

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 χ before all other alternatives $i \neq j$:

$$PDF_j(t) = f_j(t)\prod_{i \neq j}(1 - F_i(t)) \tag{4.9}$$

Thus we have:

$$Prob(j) = \int_0^\infty PDF_j(t)dt. \tag{4.10}$$

---

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

[6]For full derivations of equations 4.7, 4.8 and 4.9, 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 4.1. 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 4.1: Impact of changing the unit of time on attribute importance estimates

		Cost	Time
	original weight	0.600	0.400
DFT coefficients	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], β_{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 calculated as:

$$V_t = C \cdot M^* \cdot W_t + \varepsilon_t \tag{4.11}$$

where M^* is the original attribute matrix M but with each attribute multiplied by its corresponding element from the vector β_{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 Chapter 2. 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,

---

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

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. Chapter 2 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 4.2 as

$$d_j = \delta_j + v_j + I_0, \tag{4.12}$$

where δ_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 δ_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 λ_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', β_k (an approach analogous to the attribute-specific scaling coefficients that we defined for DFT) in Equation 4.4, which becomes:

$$w_{x_{k,i,j}} = exp(-\lambda \cdot \beta_k \cdot |x_{k,i} - x_{k,j}|) \tag{4.13}$$

However, we can relax the limitations of attribute bias and directionality simultaneously by also making an adjustment to the value function (Equation 4.3), which now takes the same form as that of the original specification with the exception that we add in attribute-specific scaling coefficients, β_k. This results in the value function from Equation 4.3 being redefined as:

$$v_j = \sum_{j \neq i} \sum_{k=1}^{K} (w_{x_{k,i,j}} \cdot \beta_k \cdot (x_{k,j} - x_{k,i})). \tag{4.14}$$

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 (ϕ_1 and ϕ_2, the sensitivity and memory parameters respectively, the number of timesteps, t, and the variance of the error term, σ_ϵ^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 χ, 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 λ_1 and λ_2).

---

[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.

2. Methodology: contrasting and improving models from mathematical psychology

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 4.2, 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 2.2) is drawn from a normal distribution with mean 0 and variance σ_ϵ^2. Consequently, as $\sigma_\epsilon^2 \geq 0$, we instead estimate the standard deviation, σ_ϵ. 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, ϕ_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]). Additionally, for choice-only data (i.e. where no additional process information is available), the start and threshold parameters A and χ are perfectly confounded. Indeed, multiplying A and χ 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 χ. The threshold χ 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 dou-

Table 4.2: 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	ϕ_1	sensitivity	> 0	ϕ_1^*	$\phi_1 = exp(\phi_1^*)$
MLBA	λ_1	similarity	≥ 0	λ_1^*	$\lambda_1 = exp(\lambda_1^*)$
MLBA	λ_2	similarity	≥ 0	λ_2^*	$\lambda_2 = exp(\lambda_2^*)$
MLBA	I_0	drift rate constant	≥ 0	I_0^*	$I_0 = exp(I_0^*)$
MLBA	A	start	fixed	not estimated	n/a
MLBA	χ	threshold	$\geq A$	χ^*	$\chi = A*(1 + exp(\chi^*))$
MLBA	s	drift rate standard deviation	fixed	not estimated	n/a

bled[9]. Finally, an MLBA model which does not find a similarity effect will result in λ 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 β_x is doubled, the positive constant I_0 is doubled, and the value of each similarity parameter λ_1 and λ_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 4.3 shows the results for the first SC dataset. Where appropriate, we used the constraints from Table 4.2 but then report the actual transformed estimates in Table 4.3, 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 (β_{LTT}) and log-fare (β_{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, ϕ_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, ϕ_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 ϕ_1 having no impact (see Equation 2.8). As $\phi_2 = 0$ in model 1, we can thus also remove ϕ_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 accumu-

Table 4.3: 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
β_{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
β_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
β_{LTT}	est.		-1.0355					
	r. t-rat.		-2.71					
β_{LF}	est.		-1.9005					
	r. t-rat.		-5.38					
ϕ_1	est.			0.0688	0.0000			
	r. t-rat.			0.00	**fixed**			
ϕ_2	est.			0.0000	0.0000			
	r. t-rat.			0.00	**fixed**			
σ_ϵ	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			
χ	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
λ_1	est.					0.0008	0.0025	0.0000
	r. t-rat.					12.90	10.36	**fixed**
λ_2	est.					0.2068	0.0369	0.1962
	r. t-rat.					16.47	11.88	6.54

lator 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 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 χ, which is a common approach in mathematical psychology (Cataldo and Cohen, 2018; Cohen et al., 2017; Trueblood et al., 2014) (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 λ_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.

Chapter 4. An accumulation of preference: contrasts between Decision Field
Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for
travel behaviour modelling
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 4.4 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 param-
eter ϕ_2, suggesting that initial comparisons matter more than current or
more recent ones. The estimate for the off-diagonal term ϕ_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 λ_2 compared to λ_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).

---

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

Table 4.4: 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**
β_{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
β_{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
β_{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
β_{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
β_{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
β_{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
β_{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
β_{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
ϕ_1	est.			0.0205	0.0000			
	r. t-rat.			1.17	**fixed**			
ϕ_2	est.			-0.5586	0.0000			
	r. t-rat.			-4.90	**fixed**			
σ_ϵ	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			
χ	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**
λ_1	est.					0.1188	0.3490	0.1191
	r. t-rat.					0.33	7.12	2.03
λ_2	est.					1.0805	2.0307	1.0556
	r. t-rat.					0.39	16.42	4.33

Chapter 4. An accumulation of preference: contrasts between Decision Field
Theory and the Multi-attribute Linear Ballistic Accumulator and adaptations for
travel behaviour modelling
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, 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 pur-
pose (employees' business, commute (TT_C in Table 4.5) 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 pa-
rameters are estimated (travel cost, TC, and headway, HW). For all three
attributes, log values are used (Arup, ITS Leeds and Accent, 2015). Addi-
tionally, Arup, ITS Leeds and Accent (2015) use three alternative specific
constants for train services run by Chiltern railways (ASC_C), Midlands rail-
ways (ASC_M) and Northern rail (ASC_N). Finally, two parameters are in-
corporated to capture income effects. Value of travel time coefficients (β_{TT_n})
are calculated for each individual n:

$$\beta_{TT_n} = \beta_{TT_{i,n}} \cdot \left(RI_n{}^{\lambda_{inc}} \cdot (1 - z_{miss,n}) + \lambda_{miss} \cdot z_{miss,n} \right) \qquad (4.15)$$

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, λ_{inc} is an income elas-
ticity on the time sensitivity and λ_{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 4.5 provides model
estimates for these parameters under MNL, RRM, DFT and MLBA.

---

[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 4.5: 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
λ_{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
λ_{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
ϕ_1	est.			1.1530	0.0000			
	rob. t-rat.			0.64	**fixed**			
ϕ_2	est.			-0.0865	0.0000			
	rob. t-rat.			-1.71	**fixed**			
σ_ϵ	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			
χ	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**
λ_1	est.					0.1107	0.1107	0.1073
	rob. t-rat.					11.20	7.96	12.12
λ_2	est.					4018.6448	Inf	Inf
	rob. t-rat.					0.76	**fixed**	**fixed**

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 λ_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 χ results in an insignificant impact on model fit with a lower BIC value obtained for model 3 compared to model 2.

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 4.6 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 4.6: 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 4.7 gives the log-likelihoods of the estimation and validation subsets of the data under each model. Additionally, Figure 4.2 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

Table 4.7: 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

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 4.8 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.

Table 4.8: 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

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
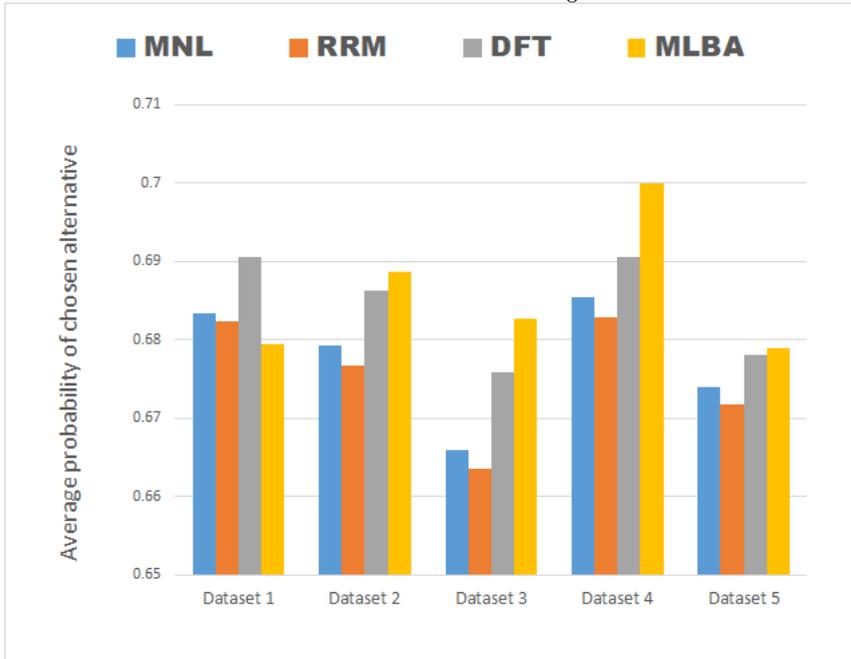
Fig. 4.2: Average probabilities of the chosen alternatives for each holdout subset of the RP data.

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

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.

Table 4.9: 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

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 Chapter 2 for an illustration of the results of failing to do this for DFT models)

3. Standard score normalisation, as previously found to be effective for DFT (see results in Chapter 2).

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

---

[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.

Table 4.10: 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**

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:

$$U_{jnt} = ASC_j + ASC_{F_j} \cdot z_{F,n} + \beta_{TT} \cdot \alpha_{TT_j} \cdot TT_{jnt} + \beta_{TC} \cdot TC_{jnt} \cdot \alpha_{IE,n} + \beta_{AT} \cdot AT_{jnt} + \epsilon_{jnt} \tag{4.16}$$

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 β_{TT} is

estimated, with appropriate adjustments applied by multiplying by α_{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 α_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 Chapter 2). 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 (δ_j) by adding them to the mean drift rates as in Equation 4.12. All of the values used for the parameters to generate probabilities for each alternative are given in Table 4.12.

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: β_{AT}, β_{TC} and a single coefficient for travel times across all alternatives, β_{TT}. Additionally we have a parameter for income effects, α^I.

2. The basic model with three additional mode-specific travel time coefficient multipliers, α_{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 4.11, with a plot of these values given in Figure 4.3. For all of the MLBA models in this section, we find that fixing χ has no significant impact on model fit.

Table 4.11: 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**

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 4.11). 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 4.3). 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.

Fig. 4.3: BIC values of the models for the simulated data

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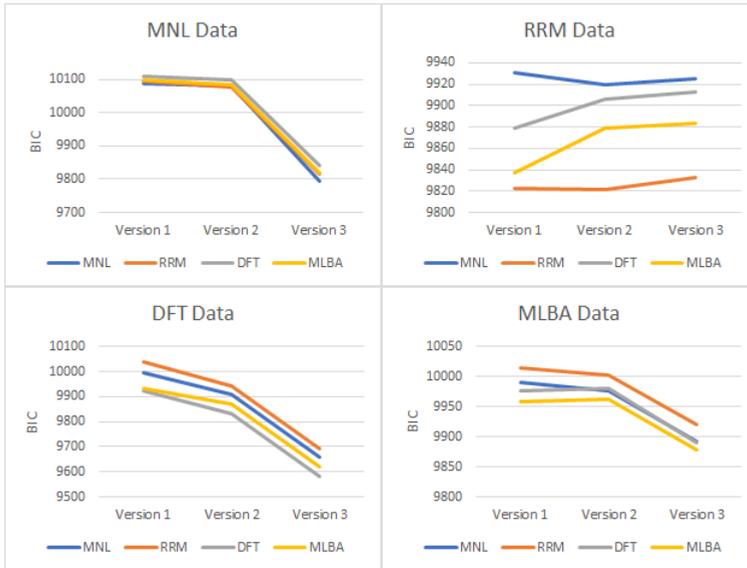
[13]It is also worth noting that this is the only dataset in the paper in which RRM significantly outperforms other models. It achieves worse fit than both DFT and MLBA on all SP and RP datasets.

[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.

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 4.4 and show that DFT 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.

Fig. 4.4: Log-likelihood of estimated models compared to model consistent with data generating process (DGP)
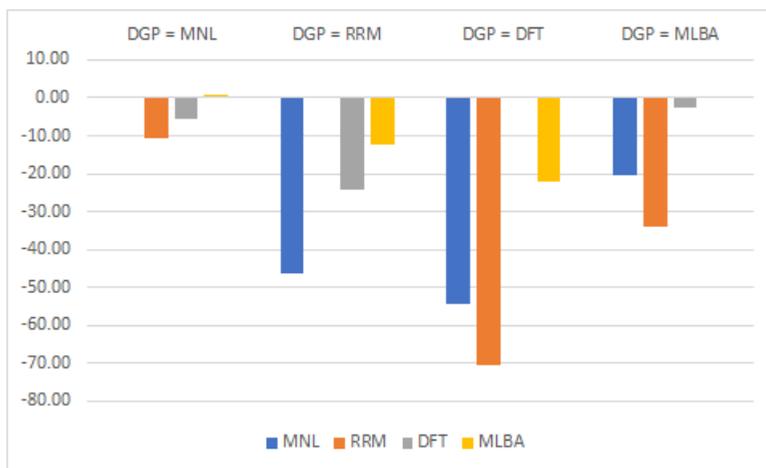
4.3 Recovery of parameters from simulated datasets

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 4.12 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 β-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.

Table 4.12: 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
β_{TT}	-0.0050	-0.0044	-12%	-0.0030	-0.0029	-3%	-0.0050	-0.0050	**fixed**	-0.0030	-0.0023	-23%
β_{TC}	-0.0280	-0.0279	0%	-0.0160	-0.0162	1%	-0.0280	-0.0308	10%	-0.0160	-0.0157	-2%
β_{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	1.1109	-113%	1.0000	1.1815	18%
$\beta_{TT_{air}}$	1.2500	1.1041	-12%	1.2500	1.7576	41%	1.2500	-0.1289	-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%
α_I	-0.5000	-0.5106	2%	-0.5000	-0.4962	-1%	-0.5000	-0.3584	-28%	-0.5000	-0.5218	4%
ϕ_1							0.0500	0.0356	-29%			
ϕ_2							0.1000	0.1379	38%			
σ_ϵ							1.4142	1.4314	1%			
t							10.0000	8.8319	-12%			
A										2.5000	2.5000	**fixed**
χ										7.5000	6.3023	-16%
s										2.0000	2.0000	**fixed**
I_0										10.0000	9.5697	-4%
λ_1										0.1000	0.0872	-13%
λ_2										0.2000	0.1702	-15%

122

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.

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 χ, 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 param-

eters. 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 feedforward 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 (Burda et al., 2008; Dumont et al., 2015; Train, 2001).

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 Chapter 2 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

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.

References

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).

References

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.

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.

References

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.

References

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.

References

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.

Chapter 5

Quantum probability models: a new framework for modelling choices

Thomas Hancock[1], Stephane Hess[1], & Charisma Choudhury[1]

Abstract

There has been an increasing effort to increase the behavioural realism of the mathematical models of choice, resulting in efforts to move away from random utility maximisation (RUM) models. Some new insights have been generated with, for example, models based random regret minimisation (RRM). However, many of the alternatives to RUM tested on real-world data, have looked at only modest departures from RUM, and differences in results have consequently been small. In this chapter, we address this research gap by investigating the applicability of models based on quantum theory - which are substantially different from the state-of-the-art choice modelling techniques. These models emphasise the importance of contextual effects, state dependence and the impact of choice or question order. We consider how to best operationalise quantum probability into a choice model. Two of our specifications find good model fit across three route choice datasets. Additionally, we test the quantum model frameworks on a best/worst route choice dataset and demonstrate that they find useful transformations to capture differences between the attributes important in a favourite alternative compared to that of the least favourite alternative. We additionally find that these 'quantum rotations' can be used to efficiently capture contextual effects where the order of the attributes and alternatives are manipulated, moral choice behaviour in the context of the addition of a taboo trade-off and can explain the difference in making trade-offs affecting just the decision-maker compared to trade-offs involving both the decision-maker and their partner. Overall, it appears that

---

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)

models incorporating quantum concepts hold significant promise in improving the state-of-the-art travel choice modelling paradigm through their adaptability and efficient modelling of contextual changes.

1   Introduction

Random utility maximisation (RUM) framework has dominated the travel choice modelling field for many decades. More recently, RUM has been criticised as being inadequate in explaining the full range of behavioural complexity (Chorus et al., 2008; Guevara and Fukushi, 2016). This has resulted in many attempts to better incorporate behavioural concepts into travel behaviour models, including regret (Chorus, 2010; Chorus et al., 2008), contextual relative advantages (Leong and Hensher, 2014) and prospect theory (Avineri and Bovy, 2008). However, none of these developments have yet rivalled RUM as the preferred model in real-world applications. This is due to difficulties that quickly arise once a modeller departs from the firm economic foundations of RUM (Hess et al., 2018). Consequently, caution is required if we are to step away from random utility models. Departures to models with similar underlying structure, such as random regret minimisation (Chorus, 2010; Chorus et al., 2008), which have the same error structure, result in only small differences whilst facing the same key fault of all departures from RUM, the loss of the ability to calculate welfare measures. Departures to more different models, such as decision field theory (Busemeyer and Townsend, 1992), whilst sometimes finding improvements in model fit, additionally result in models that may become computationally infeasible for large-scale datasets (see Table 2.7 in Chapter 2). Thus, if we are to move away from RUM, we need to investigate alternative approaches that are computationally simpler - yet, better reflect behavioural realism. This leads us to explore ideas from other disciplines which are further away from the tried and tested. Given the success of using ideas from quantum physics in cognitive psychology, one possible alternative is to see if quantum physics can make a similar step into travel behaviour modelling.

Quantum physics, first considered in the early 20th century, was originally created to explain phenomena and results that could not be explained by classical theories of probability and physics. In particular, physicists noticed that the measurement of one variable could impact the measurement of another. The most famous example of this relates to measuring the position and momentum (mass multiplied by velocity) of a particle. Physicists found that they could not measure both accurately at the same time. This led to 'Heisenberg's uncertainty principle' (Heisenberg, 1927). Formally, this could

be written (Kennard, 1927) as:

$$\sigma_x \cdot \sigma_\rho \geq \frac{\hbar}{2}, \tag{5.1}$$

where σ_x and σ_ρ are the standard deviations of position and momentum respectively, which multiplied together give the uncertainty, and $\hbar$ is the reduced Planck constant, $\frac{h}{2\pi}$ (with h the Planck constant, a physical constant first used by Planck (1901) that relates the energy carried by a photon to its frequency). To illustrate how this led to the breakdown of classical probability, we first imagine that we have three possible propositional variables:

A : 'the particle has momentum in the interval $[\rho_1,\rho_2]$'

B : 'the particle is in the interval $[x_1,x_2]$'

C : 'the particle is in the interval $[x_2,x_3]$.'

If we have a system where $\hbar = 1$, the minimum allowed uncertainty by Heisenberg's uncertainty principle (Equation 5.1) is $\frac{1}{2}$. Under this system, we might observe a particle such that the propositional variables (A) and (B or C) are true (where total uncertainty is greater than $\frac{1}{2}$). Simultaneously, it is possible that both (A and B) and (A and C) each have total uncertainties of less than $\frac{1}{2}$ and are thus false. Consequently, the distributivity law of classical probability ($A(B + C) = AB + AC$) fails to hold.

This resulted in the creation of a new theory of probability, known as quantum logic (Birkhoff and Von Neumann, 1936). Under quantum logic (which is also known as quantum probability), a new set of probability rules were defined, which crucially did not include the axiom of distributivity. This new theory of probability has subsequently made the transition into cognitive psychology (Bruza et al., 2015) and has also been introduced into transport behaviour modelling. For example, Vitetta (2016) introduced a quantum model based on random utility models with the addition of an interference term for route choice problems. Additionally, Yu and Jayakrishnan (2018) demonstrated that quantum cognition models can efficiently be used to capture the difference in state of mind between choices made under stated preference and revealed preference settings. However, thus far, as far as the authors are aware, there has not been a choice model developed with quantum concepts that incorporates attribute values for individual alternatives and can work for general choices as well as 'changes in perspective'. Thus the focus of this chapter is to explore ways to develop a choice modelling framework based on quantum logic that can be used for choices in general, as well as efficiently capturing effects caused by ordering, some form of interference or some change in 'state of mind'.

The rest of this chapter is organised as follows. First, we introduce quantum logic and discuss the relative benefits of using such a system. We then

mathematically describe quantum logic, giving detailed graphical examples. Next, we discuss how it can be incorporated into a choice model, detailing two different formulations for new models. We then test the performance of our proposed models against typical choice models such as multinomial logit and also random regret minimisation in the context of travel decisions. Finally, we test the models on best-worst and contextual choice data, before drawing some conclusions.

2 Quantum logic

In this section, we first give a general overview of quantum logic as well as detailing why and how it has been used in the past. We then give the mathematical definitions for how quantum logic works for basic choices. We conclude by describing how it works for a series of related choices. It is in the transformation from one choice task to another that a modelling framework based on quantum logic looks very different to traditional choice models.

2.1 Overview of quantum logic

A simple example of how quantum logic works is given in Figure 5.1. Initially, a decision-maker might be making a single choice between two alternatives, travelling by car or by train. Each of these alternatives are represented by vectors, $|T\rangle$ and $|C\rangle$ respectively (the axes in the Figure 5.1). Under quantum logic, the decision-maker has some initial state, denoted $|z\rangle$, regarding whether they will choose car or train.

The action of making a choice (or equivalently making a judgement or coming to some result) results in a 'change of state'. This can be represented graphically by moving from the initial state vector and 'projecting' onto the vector corresponding to the chosen alternative. In this example, ρT, represents the scalar projection of $|z\rangle$ onto a straight line parallel to $|T\rangle$. The length of this projection is then denoted $|\rho T|$. In Figure 5.1, these projections are directly over the corresponding vectors. The arrows from the initial state to these vectors are not the projections themselves, but help to determine the length of the projections, which are the distances from the origin to the points where the arrows meet the vector. When the state vector is at 45 degrees, the projections are of equal length and the probabilities are thus 50% each. In the example in Figure 5.1, the car alternative has a higher probability. The full mathematical description for this is given in the following section on a basic choice under quantum logic, which also gives a 3-dimensional example. The longer the projection onto the vector for an alternative, the more likely that alternative is chosen. The crucial difference in using such a system is how an additional question or nudge can then

SINGLE QUESTION
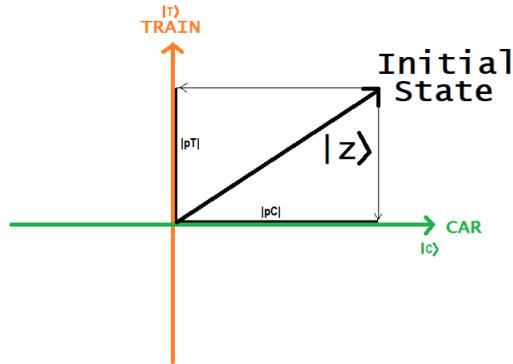
Fig. 5.1: A single question under quantum probability

impact the decision-maker's choice for the first question (car or train). If, for example, the decision-maker was asked 'are you environmentally friendly?' before they had made up their mind between the choice of car or train, they would then be initially answering a different question and making a different choice (see Figure 5.2).

As a result of the decision-maker deciding 'I am environmentally friendly', the decision-maker's state moves from the initial starting state and is projected onto the vector representing 'environmentally friendly' and vice versa if they decide 'I am not environmentally friendly' (see Figure 5.2). This results in making the choice between car and train from a different state. Consequently the length of the projections ($|\rho C|$ and $|\rho T|$) onto the vectors for car and train have changed. This is graphically represented in Figure 5.2, with the projection length for $|\rho T|$ longer if the initial state is first projected onto the environmentally friendly vector before being projected onto the train vector, relative to the projection length if train is chosen directly from the initial state. Consequently, the probabilities for choosing car and train are altered.

2.2 Key reasons for using quantum logic

Cognitive psychologists have many key reasons for using quantum logic (Busemeyer et al., 2011) that are also relevant for transport behaviour modelling. Firstly, a behavioural state is initially 'indefinite' and is often created rather than just recorded by an attempt to measure it. For example, a decision-

ADDITIONAL QUESTION
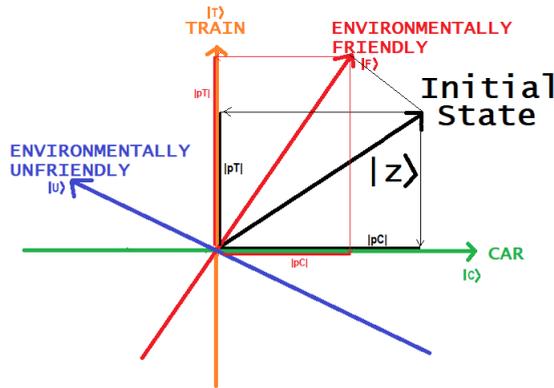
Fig. 5.2: Making two choices under quantum logic

maker might only start considering how environmentally friendly they are after they have been asked (or reminded) about how environmentally friendly they are. For this reason, it is essential that surveys including both choice tasks and attitudinal questions require the respondent to complete the choice tasks first, if the researcher wishes to avoid bias in the choice task (Ben-Akiva et al., 2019). However, conversely, a decision-maker may try to 'justify' their choices with their responses to the attitudinal questions (Cunha-e Sá et al., 2012). Consequently, it is often difficult to measure a decision-maker's *true* attitudes, opinions and preference without some form of bias. It is easy to see how this relates to issues for choice modellers with, for example, analysts often having concerns about the biases or truthfulness within stated preference data (Mahieu et al., 2016). Secondly, psychologists have put forward the argument that cognition behaves like a wave rather than a particle (Trueblood and Busemeyer, 2012). A decision-maker might consider the advantages of getting the train but then also consider the advantages of driving. Indeed, many models developed in mathematical psychology assume preferences for alternatives that update stochastically (Busemeyer and Townsend, 1993; Krajbich et al., 2012). Under quantum logic, their preference over time 'behaves like a wave' and consequently fluctuates over time. It is only when a decision-maker makes up their mind that their preference exists as some measurable definite state. Before an action or choice is made, an observer does not know what the decision-maker will do. There are many preference states within travel behaviour that could similarly be described as 'wave-like', such as anticipating merging onto a new lane when driving, changing travel

mode when weather worsens, or choosing which route to take depending on
traffic conditions. One of the most crucial quantum concepts, however, is
the idea of interferences or nudges (such as the previous example of being
asked about the environment whilst in the process of making a mode choice).
After the development of quantum physics to explain ordering effects of ob-
served variables (Birkhoff and Von Neumann, 1936), a wide range of quantum
models, often based on the idea of quantum interference, have been put for-
ward in cognitive psychology (Bruza et al., 2015). These include a quantum
model to explain ordering effects (Trueblood and Busemeyer, 2011), a quan-
tum similarity model (Pothos and Busemeyer, 2013), a quantum judgement
model (Busemeyer et al., 2011) and the use of quantum models to explain
violations of the 'sure thing principle' (Pothos and Busemeyer, 2009)[2].

These models perform a similar function to choice models that include
state dependence, where a number of different models (Seetharaman, 2003)
have been applied to capture the temporal correlation of choices over time.
However, most crucially, the adoption of quantum logic allows for an ele-
gant and convenient framework for understanding these 'paradoxical' findings
which become 'intuitive' (Wang et al., 2013).

Given the success of quantum models at explaining ordering effects within
cognitive psychology, there is ample scope for quantum logic and quantum
ideas within travel behaviour modelling and choice modelling in general. In
particular, it may also be of key interest to choice modellers studying moral
choice behaviour as the adoption of such methods 'allows for a re-examination
of the challenge of formalising psychological concepts of conflict, ambiguity,
and uncertainty' (Wang et al., 2013).

Moral choice situations can be summarised as those where a decision
maker feels that the choice alternatives can to some extent be categorised as
'right' or 'wrong'. As a result, the associated choices can be perhaps more
complex as they do not involve straightforward trade-offs between attributes
of alternatives. For example, a decision-maker may not choose what they
want to choose as they believe it to be an immoral option. While moral
choice behaviour has received much attention in economics and psychology,
it is rarely considered in the choice modelling literature (Chorus, 2015). This
is despite the fact that many typical experiments conducted for understanding
moral preferences use paradigms such as variations of the infamous trolley
problem (where a 'runaway trolley' has two possible paths, both of which will
result in the death of some individual(s), and the decision-maker must choose

---

[2]A classic example of a probability judgement error is given by Tversky and Kahneman
(1983), who found that participants, after reading 'Linda was a philosophy major. She
is bright and concerned with issues of discrimination and social justice', were more likely
to agree with the statement 'Linda is a feminist bank teller' than the statement 'Linda
is a bank teller'. The 'sure thing principle' states than an individual who would take the
same action if some event happens or not should also take that action without knowing
the outcome of the event (Savage, 1954).

who to save), for which a precise understanding of the trade-offs that are being made could be obtained using choice models. This is perhaps due to the fact that moral preferences are difficult to investigate outside of the laboratory, with typical experimental methods for examining moral choice scenarios often suffering from low external validity (Bauman et al., 2014). However, more recently, moral choice behaviour has become more prominent to the travel behaviour modelling community through, for example, the reinvention of the trolley problem as a self-driving car problem (Awad et al., 2018). Thus far, there has not been much consideration given to the types of choice models used for the modelling of such scenarios, despite the wide range of theoretical explanations for moral behaviour that have been proposed (Chorus, 2015). Whilst some steps towards the development of choice models specifically for moral choice contexts have been made (Chorus et al., 2018), an additional aim of this chapter is thus to test whether quantum logic can also be used to accurately capture changes in choice context within moral choice scenarios.

2.3   A basic choice under quantum logic

Under quantum logic, a measurement (or choice scenario), X, can be represented geometrically as a subspace L_x in a multidimensional Hilbert space (Trueblood et al., 2014b). For each measurement, a number of discrete 'events' are possible. These events, if mutually exclusive, are represented by orthonormal vectors, which are denoted $|x_1\rangle$, $|x_2\rangle$, ... $|x_J\rangle$ (with J the number of alternatives). For these vectors, we use bra-ket notation in keeping with the standard notation used in quantum mechanics and quantum cognition (c.f. Trueblood and Busemeyer 2011). Under bra-ket notation, a column vector in a Hilbert space is represented by a 'ket' vector, $|\cdot\rangle$, with the corresponding row vector (with each element being complex conjugated) a 'bra' vector, $\langle\cdot|$ (Yu and Jayakrishnan, 2018). These orthonormal vectors then form a basis for the subspace L_x. Consequently, the Hilbert space for a choice task with J alternatives can be represented by a J-dimensional space. This means that for a choice set where there are three alternatives, the Hilbert space is a 3-dimensional space and can be represented as shown in Figure 5.3.

Under quantum logic, a decision-maker has some opinion, initial state or 'indefinite state', denoted $|z\rangle$, which can be represented by a vector of unit length (see Figure 5.3). When a decision-maker makes a choice, their state goes from 'indefinite' to 'definite', by projecting onto the vector representing the chosen alternative. This means that for each event alternative L_{x_i}, there is a corresponding projection operator ρ_{x_i} that projects $|z\rangle$ onto the vector $|x_i\rangle$.

As the indefinite state vector is of unit length and the subspace is represented by a set of orthonormal vectors, the sum of the squared length of the
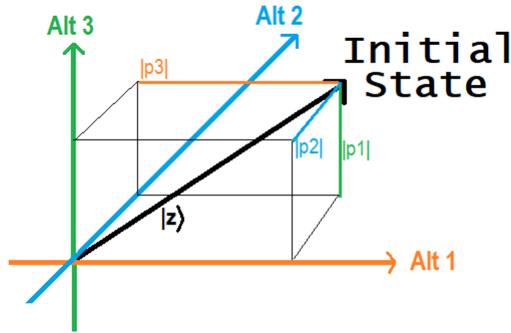
Fig. 5.3: A basic example of 3-dimensional Hilbert probability space

projections must sum to 1:

$$\sum_{i=1}^{J} |\rho_{x_i}|^2 = 1. \tag{5.2}$$

A visual proof of this fact is given in Figure 5.3. The lengths of the three projections can be visualised as the three sides of the cuboid in 3-dimensional space. By Pythagoras' theorem, the fact that the vector of unit length cuts diagonally from one corner to the opposite corner of the cuboid means that the squared lengths of the projections must sum to one.

2.4 A sequence of choices

If a decision-maker makes a second choice across a different set of alternatives, this choice may be influenced by the first. Quantum logic captures this by representing the two events by two separate subspaces within the Hilbert space, L_x and L_y. Each subspace is separately defined by the set of orthonormal vectors representing the alternatives for each event. This means that L_x is spanned by $|x_1\rangle$, $|x_2\rangle$, ... $|x_J\rangle$ and L_y is spanned by $|y_1\rangle$, $|y_2\rangle$, ... $|y_K\rangle$, where there are J alternatives for choice scenario X and K alternatives for scenario Y.

Revisiting the example presented in Figure 5.2, a decision-maker might be initially making a choice, X, between commuting by car or train. Under quantum logic, the decision-maker has some initial state (which could be based on past experiences) regarding whether they will choose car or train. All possible states are spanned by the basis vectors $|x_{car}\rangle$, $|x_{train}\rangle$. The closer the vector representing the decision-maker's state is to the vector representing an alternative, the more likely that alternative will be chosen. However, the decision-maker could first be asked some a different question

(Y) about whether they consider themselves to be environmentally friendly or not. Under quantum logic, the 'indefinite state' does not change. This means that if the probabilities for alternatives being chosen in question Y are to be different from the probabilities for alternatives being chosen in question X, we need choice Y to be represented[3] by a different set of basis vectors, $|y_{env-friendly}\rangle$, $|y_{env-unfriendly}\rangle$. Consequently, if the decision-maker makes the choice 'I am environmentally friendly', they move through the Hilbert space and their state is projected onto the environmentally friendly vector, $|y_{env-friendly}\rangle$ (see Figure 5.2). This means that their new state is the vector $|y_{env-friendly}\rangle$ itself. By making choice (Y) first, the original choice X between car and train is made from a different state (or perspective).

Crucially, by moving state (and making what we define a 'quantum rotation'), the squared lengths of the projections onto the vectors for train and for car have changed[4]. As a result, in this example, the decision-maker is more likely to choose to commute by train if they first decide that they are environmentally friendly. This is graphically represented in Figure 5.2, where the length of the projection onto the vector representing train being chosen has increased (which results in the probability of choosing train increasing relative to the probability of choosing car).

3 Building a choice model from quantum logic

Whilst Lipovetsky (2018) has applied quantum models to consumer recall tasks with multi-alternative, multi-attribute alternatives, quantum probability has not ever been applied to multi-alternative, multi-attribute choice scenarios (as far as the authors are aware). In this section, we look at how we can use ideas from quantum probability within a choice model. We do this by first considering what the requirements are for a quantum choice model. We then explore the use of sine and cosine functions, a method which was previously found to be effective for multialternative scenarios (Lipovetsky, 2018). Next, we consider ideas from two different existing choice models to use within a quantum model framework: regret functions from random regret minimisation (Chorus, 2010) and drift rate functions from the multi-attribute linear ballistic accumulator model (Trueblood et al., 2014a). Finally, we con-

---

[3]It is important here to not visualise the state as a point (a, b). It is a vector, $\left|\begin{bmatrix} a \\ b \end{bmatrix}\right\rangle$, meaning that the projection lengths from this state to the set of basis vectors will change if we use a different set of basis vectors.

[4]Two choices that require a different set of basis vectors are known as 'incompatible'. If the choices are in fact compatible and can be represented by the same set of basis vectors, then the order in which the choices are made has no impact on the probabilities of each alternative being chosen. Consequently, quantum probability collapses back into classical probability (Hughes, 1992).

sider how similar or related choices could be mathematically explained by a 'quantum rotation'.

3.1 Requirements

For our choice model to use quantum logic, we need to define a method for calculating an indefinite state vector. If this state vector is of unit length and we take projections from it to a set of orthonormal basis vectors (with one vector for each discrete alternative), then the sum of the squared length of these projections will equal one. Consequently, for each alternative, we need to find the length of the projection, as the square of this length equals the probability with which the alternative is chosen (see Figure 5.3). This means that if we are to use quantum logic to understand multi-attribute, multi-alternative choices, we must first consider how best to represent the state vector, $|z\rangle$. If, for example, we imagine that we are making a route choice between a cheap and slow alternative and a fast and expensive alternative, the development of a state could be represented by Figure 5.4.
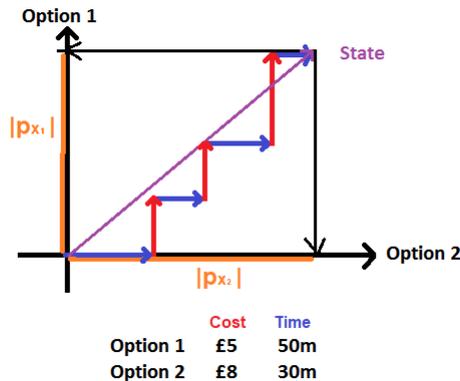
Fig. 5.4: The development of a 'state'

When the decision-maker considers travel time, the second option is better, and consequently the state vector extends in the direction of the vector representing option 2 (and hence increasing the length of the projection onto option 2 whilst not increasing the length of the projection onto option 1). If the decision-maker considers cost, the state vector instead extends in the direction of the vector for alternative 1. At some point the decision-maker makes the choice when they reach some state. To generate this state, we need to know the relative importance of the attributes. This means that one option is to calculate 'utilities' or 'preference values' for each alternative. However, if we used a utility specification from multinomial logit, $U_j = \beta' x_j$, where β is a vector of coefficients and x_j is a vector of observed variables relating

to alternative j, then some alternatives could have positive utilities and others might have negative utilities. As the probability of an alternative is the squared length of the projection from the state vector onto the vector for the alternative, positive and negative values would lead to the same result. This means that we require a method for comparing attributes across alternatives that always results in positive (or always negative) values. Additionally, a quantum choice model could have parameters equivalent to alternative specific constants, which would estimate the starting point for the state (which may not necessarily be zero). This in effect acts as an initial preference state in the multidimensional space. Finally, some alternatives may be compared more than others, so weights could be defined for each pair of alternatives (with one fixed for normalisation purposes).

3.2   Cosine and sine functions

It is easy to see from Figure 5.5 that cosine and sine functions provide one possibility for quantum models. As $sin^2(\theta) + cos^2(\theta) = 1$, one possible method
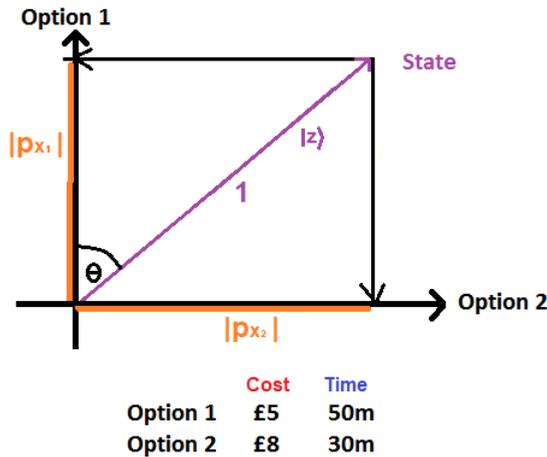
Fig. 5.5: Using geometry within a quantum model

for defining quantum probabilities is to find the angle θ between the state vector $|z\rangle$ and the projection onto the chosen alternative, ρ_{x_j}. For a choice scenario with two alternatives as defined in Figure 5.5, $cos(\theta) = |\rho_{x_1}|$ and $sin(\theta) = |\rho_{x_2}|$. Consequently, finding the angle θ will give us the probabilities of both alternatives. We thus need to define the angle θ as a weighted sum of the differences between the attributes of the alternatives. This angles can

then be used to calculate projections (ρ):

$$\theta = \delta + \sum_{k=1}^{K} \beta_k (x_{k_1} - x_{k_2}),$$
$$|\rho(Alt_1)| = sin(\theta),$$
$$|\rho(Alt_2)| = cos(\theta), \qquad (5.3)$$
$$\rightarrow \rho(Alt_1)^2 + \rho(Alt_2)^2 = 1,$$

where $delta$ is a constant, $k = 1, ...K$ is an index across the attributes, β_k is a coefficient for attribute k and x_{k_i} is the value of attribute k for alternative i. This approach is used by Lipovetsky (2018) to predict which pizza brands participants remember. He demonstrates that for more than two alternatives, we simply need to find additional angles:

$$[sin^2(\theta_A)] + [cos^2(\theta_A)] = 1$$
$$\rightarrow sin^2(\theta_A) + cos^2(\theta_A)(sin^2(\theta_B) + cos^2(\theta_B)) = 1$$
$$\rightarrow [sin^2(\theta_A)] + [cos^2(\theta_A) \cdot sin^2(\theta_B)] + [cos^2(\theta_A) \cdot cos^2(\theta_B)] = 1 \qquad (5.4)$$
$$\rightarrow [sin^2(\theta_A)] + [cos^2(\theta_A) \cdot sin^2(\theta_B)] +$$
$$[cos^2(\theta_A) \cdot cos^2(\theta_B) \cdot sin^2(\theta_C)] + [cos^2(\theta_A) \cdot cos^2(\theta_B) \cdot cos^2(\theta_C)] = 1.$$

However, when there are more than two choice alternatives, complications arise. Firstly, it is not clear which differences between alternatives should be used, as we only require $J - 1$ differences where J is the number of alternatives. Crucially, under this system, a small increase in the attribute of one alternative may not have the equivalent impact as a small increase in the attribute of a different alternative. An example of this is given in Table 5.1. In this set of examples, we have three alternatives {A,B,C} and

Table 5.1: Examples of probabilities under various coefficients for sine and cosine models

	Parameter/Attribute	Ex. 1 coefs.	Ex. 2 coefs.	Ex. 3 coeffs.
Constant A:	x	0.615	0.615	0.615
Constant B:	y	0.785	0.785	0.785
Attribute for Alt A:	Z_A	3	3.1	3
Attribute for Alt B:	Z_B	3	3	3.1
Attribute for Alt C:	Z_C	3	3	3
Alternative	Projection length	Ex. 1 prob.	Ex. 2 prob.	Ex. 3 prob.
A	$sin(\theta_A)$	0.333	0.430	0.243
B	$cos(\theta_A) \cdot sin(\theta_B)$	0.333	0.285	0.454
C	$cos(\theta_A) \cdot cos(\theta_B)$	0.333	0.285	0.304

just a single attribute, Z. We set the relative importance weights $\beta_k = 1$, $\theta_A = x + Z_A - Z_B$ and $\theta_B = y + Z_B - Z_C$. For the first example, we demonstrate that precise constants of $x = 0.615$ and $y = 0.785$ are required to allow for the probability of the three alternatives to be identical if the

alternatives have matching value of 3 for attribute Z. If, in example 2, we imagine that the value for attribute Z_A increases to 3.1, then the probabilities adjust appropriately (the probability of alternative B remains the same as the probability of alternative C). This is not the case in example 3, where an equivalent increase in attribute Z_B does not result in alternatives A and C having the same probability. Consequently, we must look for alternative model frameworks for multi-alternative settings.

3.3   Quantum pairwise comparison framework A (QPCA)

We will now show that an alternative possibility for a quantum model framework is to use regret functions from random regret minimisation (RRM) as the key component for the definition of projection lengths for each alternative. The deterministic regret (Chorus, 2010) for respondent n, for alternative i, in choice task t, is given by:

$$R_{int} = \sum_{k=1}^{K} \sum_{j \neq i} ln(1 + e^{\beta_k(x_{jntk} - x_{intk})}) \tag{5.5}$$

with $k = 1, ..., K$ an index across attributes and β_k a coefficient for the relative importance of attribute k. This has the potential to work for a quantum model as the logarithm guarantees that only positive values are generated from the pairwise comparisons between the alternatives. We can therefore make the following definition for 'quantum pairwise comparison version A (QPCA)' to calculate the length of projection for alternative i (with n the respondent and t the choice task):

$$|\rho_{int}| = \delta_{QPCA,i} + I_0 + \sum_{k=1}^{K} \sum_{j \neq i} wt_{ij} \cdot ln(1 + e^{\beta_k(x_{intk} - x_{jntk})}), \tag{5.6}$$

where $\delta_{QPCA,i}$ are alternative-specific constants and I_0 is a constant that has the same value across all alternatives[5]. Together, these are used to define the starting point for the state vector in the Hilbert space. wt_{ij} is a weight for the relative importance of the comparison between alternatives i and j, meaning that it is constant across attributes as it is at the alternative level. Once these projection lengths have been calculated, the probability for each alternative can be defined simply as:

$$P_{QPCA,jnt} = \frac{|\rho_{jnt}|^2}{\sum_{i=1}^{J}(|\rho_{int}|^2)}, \tag{5.7}$$

---

[5]Note that x_{jntk} and x_{intk} in Equation 5.6 compared to Equation 5.5 are reversed, as using negative regret as in RRM models will be ineffective given the projection lengths are later squared.

where $i = 1, ...J$ is an index across the possible alternatives. This means that to estimate this model, we require K attribute coefficients, $\frac{(J)(J-1)}{2} - 1$ weights, wt, for the relative importance of comparisons between the different alternatives (where the use of logistic transformations ensures that these weights will sum to one) and J constants to estimate the initial starting state. Whereas adding a constant to the utility of every alternative does not have an impact in random utility models, it is multiplying all of the length of projections by a constant that does not impact the choice probabilities of alternatives under a quantum system (see Equation 5.2). Consequently, we can have J parameters for J alternatives to define the starting state (either by normalising one attribute specific constant or by not using I_0, the constant that is added to all alternatives). The greater the magnitude of these constants, the less deterministic the choices become.

3.4 Quantum pairwise comparison framework B (QPCB)

The final possibility we consider for a quantum model framework in this chapter is the use of drift rate functions from the multi-attribute linear ballistic accumulator model (Trueblood et al., 2014a). The linear ballistic accumulator (LBA), was originally designed within mathematical psychology, and is a model designed to capture both choices and response times (Brown and Heathcote, 2008). Under LBA, a decision-maker starts with a random amount of evidence for each alternative. The evidence for each alternative then grows linearly according to a set of drift rates (with one rate for each alternative). The first to reach some threshold is then the chosen alternative. This model was then adjusted for alternatives with multiple attributes (MLBA) and has been used successfully to explain choices between ratings for eyewitness testimony (Trueblood et al., 2014a), consumer and perceptual choices (Turner et al., 2018) and gambling and accommodation choices (Cohen et al., 2017). Under MLBA, the drift rates are generated from a normal distribution where the mean drift rates are a function of the attributes of the alternatives. Crucially, the mean drift rates are often defined such that they are guaranteed to be positive, meaning that they could be used within a quantum probability framework. The mean drift rate for respondent n, for alternative i, in choice task t, is defined as:

$$d_i nt = I_0 + \sum_{k=1}^{K} \sum_{j \neq i} \left(w_{x_{k,i,j}} \cdot \beta_k \cdot (x_{intk} - x_{jntk}) \right), \tag{5.8}$$

where I_0 is a constant, which can be defined such that $d_j \geq 0$ for all j. Then additionally K is the number of attributes, $w_{x_{ijntk}}$ is a similarity weighting, β_k is an attribute-specific scaling coefficient for attribute k and x_{intk} and x_{jntk} are the values for alternatives i and j for attribute k. Whilst similar in

appearance to regret functions, rather than using a logarithm, we instead use similarity weightings, which are defined such that they are an exponentially decaying function of distance:

$$w_{x_{ijntk}} = exp\Big(-(\lambda_1 \cdot [x_{intk} \geq x_{jntk}] + \lambda_2 [x_{intk} < x_{jntk}]) \cdot \beta_k \cdot |x_{intk} - x_{jntk}|\Big). \tag{5.9}$$

Under MLBA, two different values, λ_1 and λ_2, are used to capture Tversky (1977)'s findings that the subjective similarity between A and B and the subjective similarity between B and A may not be equal. Given that differences between losses and gains have regularly been shown to be important in a transport context (Hess et al., 2008; Masiero and Hensher, 2010; Stathopoulos and Hess, 2012), this is a useful feature for this quantum model as well. Both λ values should be greater than zero to ensure that attributes that are more similar have a higher similarity value $w_{x_{ijntk}}$. This results in weights that are between 0 and 1. Finally, we adjust the drift rate specification as before to include weights for the relative importance of comparisons between pairs of alternatives, and to include constants to adjust the initial state vector. Consequently, the length of the projection for an alternative j for decision-maker n in choice task t in our quantum pairwise comparison version B (QPCB) model is:

$$|\rho_{int}| = \delta_{QPCB,i} + I_0 + \sum_{k=1}^{K} \sum_{j \neq i} wt_{i,j} \cdot \big(w_{x_{ijntk}} \cdot \beta_k \cdot (x_{intk} - x_{jntk})\big), \tag{5.10}$$

where again $\delta_{QPCB,i}$ is used to define the starting point for the state vector and wt_{ij} is a weight for the relative importance of the comparison between alternatives i and j. Once these projection lengths have been calculated, we can use Equation 5.7 again to calculate the probability of alternatives under this model.

3.5 Quantum rotation

We also look at how quantum models can explain a pair of related choices through a 'quantum rotation'. Under quantum logic, a separate set of basis vectors is required for a different choice. For example, if a decision-maker is selecting their favourite of three alternatives, then the set of basis vectors could be $|Alt1_{best}\rangle$, $|Alt2_{best}\rangle$ and $|Alt3_{best}\rangle$ (see Figure 5.6 Graph A).

The decision-maker could choose any of the three alternatives and therefore there are three possible projections from the state vector $|z\rangle$ onto the three alternatives (with the corresponding projection lengths labelled $|\rho1|_{best}$, $|\rho2|_{best}$ and $|\rho3|_{best}$ in Figure 5.6 Graph A). Now suppose the decision-maker chooses alternative 2. Under quantum logic, the state changes from $|z\rangle$ to $|Alt2_{best}\rangle$. However, if the decision-maker was then asked to make a 2nd
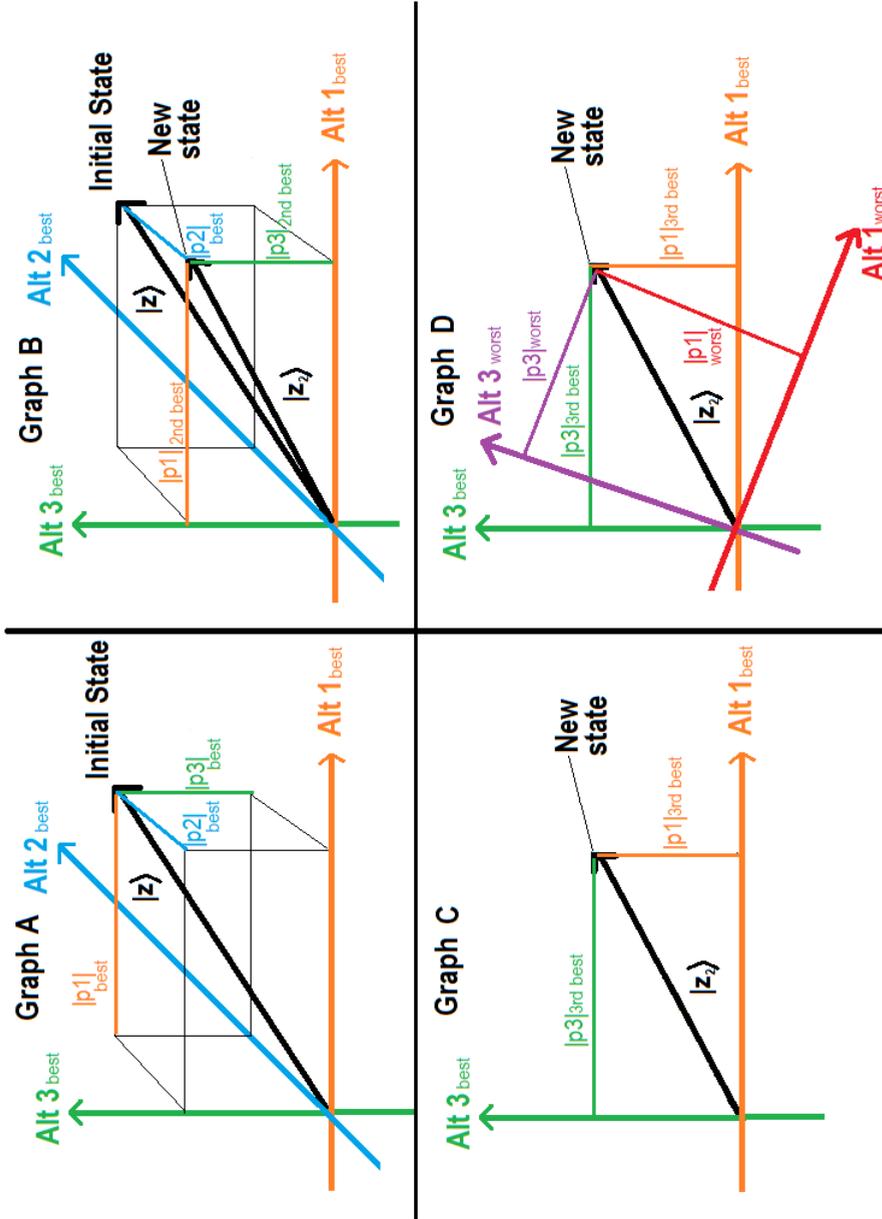
Fig. 5.6: Making best-worst choices under quantum logic

choice to pick their least favourite, then the projection lengths from $|Alt2_{best}\rangle$ to $|Alt1_{best}\rangle$ or $|Alt3_{best}\rangle$ are both zero, as these are orthogonal vectors. Consequently, to calculate probabilities for the 2nd choice, we must instead assume that the the initial state can still accurately capture the difference between the 2nd and 3rd favoured alternatives. This means the probabilities for the 2nd choice can be found by reducing the dimensionality (moving from the initial state to the new state, $|z_2\rangle$, in Figure 5.6 Graph B).

Under a typical MNL, the utility (or equivalently regret in RRM) of picking alternative i can easily be adapted to a corresponding utility for not picking alternative i:

$$U(alt_{i-worst}) = -U(alt_{i-best}). \tag{5.11}$$

For quantum models, however, this translation is not as simple. This is a consequence of using the squared length of the projections to calculate the probability of alternatives (see Equation 5.7), as using 'negative lengths' for each projection will result in the same probabilities for each alternative. However, when there are only three alternatives (a regular setting for many surveys), there exists a simple transformation. Given that there are two alternatives left, the probability of picking one option as the second best (or second most preferred) equals the probability of picking the other as the worst. Consequently given alternatives i and j, we can define the length of the projection for alternative i being the worst as:

$$|\rho_{i_{worst}}| = |\rho_{j_{best}}|. \tag{5.12}$$

This can be represented graphically by moving from Figure 5.6 Graph B to Graph C. A different translation would be required if there were more than two alternatives left. For example, one could simply use differences in favour of the alternative rather than against the alternative in the definitions of the projection length for each alternative. For example, Equation 5.6 would become:

$$|\rho_{int}| = \delta_{QPCA,i} + \sum_{k=1}^{K} \sum_{j \neq i} wt_{ij} \cdot ln(1 + e^{\beta_k(x_{jntk} - x_{intk})}). \tag{5.13}$$

Equivalently, using $x_{k,j} - x_{k,i}$ instead of $x_{k,i} - x_{k,j}$ in Equation 5.10 would result in an appropriate translation for QPCB:

$$|\rho_{int}| = \delta_{QPCB,i} + I_0 + \sum_{j \neq i} \sum_{k=1}^{K} wt_{i,j} \cdot \left(w_{x_{k,i,j}} \cdot \beta_k \cdot (x_{k,j} - x_{k,i}) \right), \tag{5.14}$$

However, Equations 5.11-5.14 only work under the assumption that 'worst' is considered directly opposite to 'best' and the sensitivities to attributes is identical when picking favoured and least favoured alternatives.

It is possible that the attributes which are important for a best choice are not necessarily the attributes which are important for a worst choice (Giergiczny et al., 2017). The simplest way for taking this into account would be to estimate a completely separate set of parameters for worst choices compared to best choices. A different possibility under quantum logic is to instead have a new set of basis vectors, $|Alt1_{worst}\rangle$, $|Alt2_{worst}\rangle$ and $|Alt3_{worst}\rangle$ representing the choice of the worst alternative. This means that if a decision-maker chose alternative 2 as the best alternative, then the probability for choosing, for example, alternative 3 as the worst, is now the squared length of the projection from $|z_2\rangle$ onto $|Alt3_{worst}\rangle$ (rather than $|Alt1_{best}\rangle$, as would be the case under Equation 5.12). This means that the length of the projection has changed (see Figure 5.6 Graph D). To calculate the new length of the projection, we would need to know how to make the change of basis from the Hilbert space represented by $|Alt1_{best}\rangle$, $|Alt2_{best}\rangle$ and $|Alt3_{best}\rangle$ to the space represented by $|Alt1_{worst}\rangle$, $|Alt2_{worst}\rangle$ and $|Alt3_{worst}\rangle$. Thus we must estimate a change of basis matrix, M^* that appropriately adjusts the lengths of the projections (with the matrix being of size n, which is the number of alternatives in the initial choice scenario). For all models, this matrix maps original utilities/lengths, p, to new ones, q, simply through matrix multiplication:

$$q_j = \sum_{i=1}^{n} M_{j,i}^* \cdot p_i \qquad (5.15)$$

Mathematically, this is effectively simultaneously a change in both underlying preferences towards alternatives and a change in scale. In our empirical application, we test this idea of a 'quantum rotation' for MNL, RRM, QPCA and QPCB to see whether the rotation can capture as much of a difference in the different sensitivities as a different set of parameters would.

4 Empirical application

In this section, we test our different specifications of quantum models on a number of route choice datasets. First, we describe the seven different datasets that we use. Three of these are basic route choice datasets, with the other four providing examples under which we can test 'quantum rotations'. We apply the models to the three basic datasets first, whilst also considering parameter estimates and out-of-sample validation for the most complex of these datasets. Next, we consider a best-worst dataset where we test the ability of our quantum models to capture both best and worst choices. We then also test quantum rotation on a dataset where the order of the alternatives and the attributes is manipulated, thus testing whether this rotation can capture ordering effects. We additionally consider out of sample validation for these datasets. Finally, we test different variations of quantum

models on our two 'moral choice' datasets. For all models, we use R packages
maxLik (Henningsen and Toomet, 2011) and apollo (Hess and Palma, 2019)
for estimation of the log-likelihood functions.

4.1   Datasets

4.1.1   Danish dataset

The first dataset we use comes from the Danish value of travel time study
(Fosgerau, 2006). 545 participants completed a total of 4,214 choice tasks.
Each task involved a simple choice between two routes, where each task has
a cheap but slow alternative and a fast but expensive alternative.

4.1.2   Swiss dataset

The second dataset we use comes from the Swiss value of time study (Ax-
hausen et al., 2008). 389 participants each make 9 binary route choice tasks.
The two alternatives are described by travel cost, travel time, headway and
the number of interchanges required to complete the trip.

4.1.3   UK dataset 1

The third dataset that we use is a survey asking public transport commuters
living in the UK to make a set of ten choices between three route alternatives
in a stated preference survey. A total of 368 participants completed the survey
resulting in 3,680 choices. Each choice task involves an invariant reference
trip and two hypothetical alternatives. Each alternative is described by seven
attributes: travel time (in minutes), fare ($\pounds$), rate of crowded trips, rate of
delays (both out of 10 trips), the average length of delays (across delayed
trips) and the availability and cost of a provision of an information service
($\pounds$). Full details of the dataset are given by Hess and Stathopoulos (2013).

4.1.4   Best-worst dataset

The best-worst dataset we use is similar to our UK dataset involving public
transport commuters. A total of 391 participants complete 10 choice tasks
described by three alternatives with the same set of attributes as before in the
previous UK dataset. As participants choose a best and a worst alternative
in each choice task, we have a total of 7,820 choices. For full details of the
dataset, readers should refer to Stathopoulos and Hess (2012).

4.1.5   UK dataset 2

The second UK dataset that we use in this chapter comes from the most recent
value of travel time study conducted in the UK (Batley et al., 2017). This

dataset comprises of 15,045 choices between two alternatives, one of which is cheaper and the other faster (SP1 in Batley et al. 2017). The advantage of this dataset is that the cheaper alternative is sometimes first and sometimes second, and additionally the order in which the attributes are displayed is also manipulated across respondents. These ordering effects have previously been found to be significant (Hess et al., 2017), making this an appropriate dataset to test quantum rotations on.

4.1.6 First moral choice dataset

The first 'moral choice' dataset we use involves 'taboo trade-offs' (Chorus et al., 2018). Decision-makers choose between the introduction of a new transport policy or keeping the status quo. To simplify the choice scenarios, each new policy offered simply an increase or decrease for four attributes: 300 EUR vehicle ownership tax, 20 minutes travel time for each car commuter per day, 100 serious injuries in traffic accidents and 5 deaths in traffic accidents. This results in a total of 16 possible new policies, which are offered in turn to each of 99 decision-makers, resulting in a dataset with a total of 1,584 choices. Chorus et al. (2018) then define a choice as involving a 'taboo trade-off' if a decision-maker could choose a policy that involves decreasing tax or travel time (a secular attribute) at the cost of increasing the number of injuries or deaths (a sacred attribute).

4.1.7 Second moral choice dataset

The final dataset we test involves decision-makers completing two sets of choice tasks based on an individual's willingness to accept longer commutes for better salaries (Beck and Hess, 2016). The first set involved trade-offs between the individual's current travel time and salary or an increased salary (of 500 or 1000 SEK in net wage per month) at a cost of an increase in one-way travel time (of either 10 or 25 minutes). The second set additionally included attributes for increased travel time and salaries for the partner of the decision-maker, meaning that the decision-maker has to make choices about who to prioritise. While the first may involve typical time-cost trade-offs that can potentially be captured well with RUM models, the latter involves a more complex decision context without any 'crisp' trade-off element in that there may not be a clear ethical protocol for how to make the decision. All choice tasks included a status quo alternative, a new alternative and a 'I am indifferent' option. A sample of of 1,179 households (with both partners in each household, resulting in 2,358 individuals) completed 4 tasks for the first set involving only attributes affecting themselves, and 4 or 5 tasks for the second set with attributes impacting both members of the household. This resulted in a total of 20,041 choice observations.

4.2 Basic models

For basic tests of our quantum models, we use the Danish, Swiss and first UK datasets. We test the models on all three datasets before testing out-of-sample performance of the models on the UK dataset.

4.2.1 Estimation results

For all datasets, we compare the quantum models against multinomial logit (MNL). We additionally test the models against random regret minimisation (RRM) for our dataset with more than two alternatives. For all of our models we use an initial parameter search algorithm, based on Bierlaire et al. (2010)'s heuristic for non-linear global optimisation. This reduces the risk of models converging to poor local optima. We test all three quantum specifications. Results for these models are given in Table 5.2.

For the quantum pairwise comparison version A (QPCA) model, we find that for the Danish dataset, there is no significant difference in model fit by including a number of different starting state vector parameters corresponding to the number of alternatives. Consequently, I_0 and δ_1 are fixed to zero. Additionally, fixing one of the attribute scaling coefficients, β_1, has no impact on model fit either. This is also the case for the Swiss dataset, but not for the UK dataset, which requires a full set of attribute scaling coefficients. Whilst the Swiss model improves with a full sets of constants, there is no significant impact on model fit for the UK dataset when all constants are removed, as the alternative comparison weights, wt, more effectively capture the underlying baseline preferences towards the different alternatives (with worse model fit obtained if alternatives specific weights are used instead of alternative comparison weights).

We observe similar results for quantum pairwise comparison version B (QPCB), which also only requires one parameter to calculate the initial starting state (δ_1 for the Danish and Swiss datasets, I_0 for the UK dataset). For identification purposes, we fix one λ parameter to a value of 1, as dividing λ by some value x and multiplying the β parameters by x results in projection lengths that are also multiplied by x (hence not changing the probability with which each alternative is chosen, see Equation 5.10). For the Danish dataset, the second λ coefficient can also be fixed to zero without impacting model fit.

Finally, we also test our trigonometric quantum model (TQ) based on the use of sines and cosines. As discussed previously, the TQ model would not work well for more than two alternatives, therefore we only test it on the Danish and Swiss datasets.

Most significantly, the QPCA and QPCB models have better model fit than both MNL and RRM across all three datasets, with a large improvement in particular for the Danish and Swiss datasets. The trigonometric quantum

Table 5.2: Results from applying quantum models to the three initial stated preference datasets

Danish				
Model	pars	LL	AIC	BIC
MNL	3	-2,301.25	4,608.50	4,627.54
QPCA	2	-2,012.25	4,028.50	4,041.19
QPCB	3	-2,010.85	4,027.70	4,046.74
TQ	3	-2,490.46	4,986.92	5,005.96
Swiss				
Model	pars	LL	AIC	BIC
MNL	5	-1,667.97	3,345.94	3,376.74
QPCA	5	-1,587.67	3,185.33	3,216.13
QPCB	6	-1,569.46	3,150.92	3,187.88
TQ	5	-1,699.16	3,408.33	3,439.13
UK				
Model	pars	LL	AIC	BIC
MNL	10	-3,360.43	6,740.86	6,802.97
RRM	10	-3,363.91	6,747.82	6,809.93
QPCA	10	-3,339.00	6,698.00	6,760.11
QPCB	12	-3,327.63	6,679.25	6,753.78

model does not perform as well, having worse fit than MNL for both the Danish and Swiss datasets. This is perhaps unsurprising given that sine and cosine function oscillate. Consequently, there is a restricted range of values that the estimated parameters can take such that the largest difference in attribute value still results in the largest value after a sine or cosine function is applied.

In Table 5.3, we also give some parameter estimates for models run on the UK data. Whilst the outputs from a quantum model cannot be translated into measures such as the value of travel time, we can get an indication of the relative importance of the attributes by dividing the parameter estimates by the sum of the absolute value of all eight attribute coefficients[6]. We find that MNL and RRM have near identical relative importances, again suggesting that they are very similar models. QPCA finds relative importances which are largely similar to their respective values under RRM, with the exception of the values for average delay and rate of delays, both of which are more important under QPCA. Under QPCB, rather different values are found. In particular, the relative importance of the fare is higher under QPCB. It is

---

[6]Note that we use a logarithmic transformation of fare as we find non-linear sensitivities in this data. In line with Hess et al. (2012), a term for reliability is also added by calculating the expected length of delay (rate of delays multiplied by average delay time).

notable that under both quantum models, the comparison of alternatives 1 and 2 is most important, followed by the comparison between 1 and 3. As alternative 1 is the reference alternative, this suggests that decision-makers give more importance to their current alternative. This implies that if they are to choose a new alternative (2 or 3) then it is more important that this alternative is better than the reference alternative than the other new alternative.

4.2.2   Validation results

We also try forecasting for the quantum models to test whether they perform well in out-of-sample validation. In this case, we split the UK dataset into five subsamples. For each subsample, we first estimate the parameters for the MNL, RRM, QPCA and QPCB models on the first 80% of the data before finding the log-likelihood of the remaining 20% validation set under the estimated parameters found for the initial set. For four and three out of five subsets respectively we observe that QPCA and QPCB have better out-of-sample log-likelihoods than MNL or RRM (see Table 5.4) assuring that the improved goodness-of-fit observed in the estimation results are not due to overfitting the data.

4.3   Models with quantum rotation

Given that quantum logic can explain ordering effects, it is also important that any quantum choice model retains this property. We therefore also look at the ability of our models to incorporate 'quantum rotations', as defined in Section 3.5.

4.3.1   Best-worst data findings

For our best-worst data, we try five different variations of each model, including three different versions of 'quantum rotations':

1. A basic structure for each model where a 'single' set of parameters are estimated for predicting both best and worst choices, where we simply take negative values (as described in Section 3.5) to translate from describing a best alternative to describing a worst alternative for MNL and RRM. For the quantum models, we try two adjustments for picking the worst alternative as oppose to the best alternative. The first method follows the rotation defined by Equation 5.12, such that the probability of choosing an alternative as second best is simply the probability that the other unchosen alternative is chosen as the worst. The second method we test is to use different projection lengths as

Table 5.3: Parameter estimates for the models ran on the UK data

UK dataset		MNL	RRM	QPCA	QPCB
Log-likelihood		-3,360.43	-3,363.91	-3,339.03	-3,327.63
AIC		6,740.86	6,747.82	6,698.05	6,679.25
BIC		6,802.97	6,809.93	6,760.16	6,753.78
β_{TT}	estimate	-0.05	-0.03	-0.39	-0.03
	robust t-ratio	-9.50	-9.58	-5.58	-2.64
	rel. importance	0.72%	0.63%	0.68%	0.48%
β_{LFare}	estimate	-6.00	-4.11	-50.29	-6.54
	robust t-ratio	-18.87	-17.66	-7.27	-2.59
	rel. importance	86.08%	86.71%	87.17%	92.70%
β_{Crowd}	estimate	-0.22	-0.15	-1.32	-0.18
	robust t-ratio	-8.58	-8.59	-4.36	-2.60
	rel. importance	3.16%	3.16%	2.29%	2.60%
β_{Delay}	estimate	-0.03	-0.02	-0.82	-0.08
	robust t-ratio	-3.25	-3.06	-2.92	-2.25
	rel. importance	0.43%	0.42%	1.42%	1.12%
β_{Rate}	estimate	-0.19	-0.12	-2.08	-0.07
	robust t-ratio	-5.96	-5.82	-5.38	-1.97
	rel. importance	2.73%	2.53%	3.60%	0.97%
β_{Rel}	estimate	-0.06	-0.04	-0.22	-0.01
	robust t-ratio	-2.64	-2.71	-2.81	-1.82
	rel. importance	0.86%	0.84%	0.38%	0.18%
β_{Inf}	estimate	-0.09	-0.05	-0.36	-0.01
	robust t-ratio	-1.13	-0.95	-0.59	-0.32
	rel. importance	1.29%	1.05%	0.62%	0.13%
β_{InfF}	estimate	0.33	0.22	2.22	0.13
	robust t-ratio	4.95	4.85	4.57	2.67
	rel. importance	4.73%	4.64%	3.84%	1.81%
asc_{alt_1}	estimate	0.39	0.27	0.00	0.00
	robust t-ratio	5.85	4.17	**fixed**	**fixed**
asc_{alt_2}	estimate	0.16	0.17	0.00	0.00
	robust t-ratio	3.3	3.38	**fixed**	**fixed**
wt_{PC12}	estimate			40.58%	43.66%
	robust t-ratio			28.90	14.14
wt_{PC13}	estimate			36.59%	36.09%
	robust t-ratio			25.29	13.31
I_0	estimate			0.00	0.24
	robust t-ratio			**fixed**	3.01
λ_2	estimate				16.17
	robust t-ratio				2.09

Table 5.4: Results from holdout samples for the different models for the UK dataset

UK dataset		MNL	RRM	QPCA	QPCB
Full Dataset	Estimate	-3,360	-3,364	-3,339	-3,328
Subset 1	Estimate	-2,651	-2,653	-2,633	-2,627
	Forecast	-713	-714	-709	-706
Subset 2	Estimate	-2,721	-2,725	-2,701	-2,688
	Forecast	-642	-641	-639	-642
Subset 3	Estimate	-2,694	-2,697	-2,670	-2,658
	Forecast	-667	-668	-670	-671
Subset 4	Estimate	-2,682	-2,685	-2,672	-2,663
	Forecast	-681	-682	-670	-668
Subset 5	Estimate	-2,684	-2,685	-2,672	-2,662
	Forecast	-679	-680	-670	-668

defined by Equations 5.13 and 5.14, which uses differences in attributes in favour of the alternative in question (rather than differences against).

2. Allowing for a completely 'separate' set of parameters for the best choices compared to the worst. This is equivalent to running two separate models where the dataset is split into two subsets: one with the only the best alternative choice tasks and one with only the worst alternative choice tasks.

3. A quantum rotation model with a diagonal rotation matrix M_1^* (with zeros off the diagonal). This results in 3 additional parameters for MNL and RRM, and 2 for the quantum models (as one needs to be fixed for identification purposes, as Equation 5.7 means that changing all projection lengths by the same constant results in the same probability with which each alternative is chosen).

4. A quantum rotation model with a symmetric rotation matrix M_2^*. This results in 5 additional parameters for all models, as one must now also be fixed for identification purposes for MNL and RRM (increasing all entries of a column of M^* has no impact on the probabilities of choosing an alternative).

5. A quantum rotation model with a fully flexible rotation matrix M_3^*. This results in 6 extra parameters for MNL and RRM and 8 extra parameters for the quantum models.

The results of these models are given in Table 5.5. For each model, the equation that is used to adjust utilities/lengths for best alternatives to

utilities/lengths for worst alternatives is given by the best-worst adjustment equations[7] at the top of Table 5.5.

Unsurprisingly, every model finds a significant improvement in model fit by having a separate set of parameters for the best alternatives compared to the worst alternatives (in line with the results of Giergiczny et al. 2017). This suggests that the relative sensitivities to the different attributes for a best alternative are not necessarily the same as the relative sensitivities to the different attributes for a worst alternative. Additionally, we also observe that the quantum models that use Equations 5.13 and 5.14 for best-worst rotations result in better model fit than models that use Equation 5.12. This is convenient in that rotations based on Equation 5.12 are inflexible and cannot be simply expanded for best-worst choice scenarios involving more than three alternatives.

Table 5.5 also gives log-likelihood gains (LL Gains) for the models using a quantum rotation as a percentage of the improvement gained by using a 'separate' set of parameters compared to a 'single' set. Crucially, we find that both quantum models (and RRM) make vast improvements by incorporating a quantum rotation, compared to models with a single set of parameters. Additionally, models with symmetric or fully flexible rotation matrices (Quantum rotation 2 and 3 in Table 5.5) provide better model fit for the second version of each quantum model than their respective 'separate' counterparts. Notably, even simple rotations work well for these models, with only two additional parameters required for QPCA and QPCB (2nd versions) to achieve 99% and 95% log-likelihood gains with respect to the improvement of using separate sets of parameters compared to a single set. This suggests that the concept of a 'quantum rotation' is particularly valid for a quantum model. It also demonstrates that the projection onto a basis vector for a best alternative is not equivalent to the projection onto a basis vector for a worst alternative. Consequently, we find that best-worst choices in this dataset are incompatible: a quantum rotation is required to move from a set of basis vectors for best choices to a different set of basis vectors for worst choices.

The impact of the 'quantum rotation' is that the probabilities of each alternative being chosen is altered. For the QPCA and QPCB (2nd version) quantum rotation models, the corresponding (fully flexible matrix) rotation matrices that are estimated are:

$$M_{QPCA} = \begin{bmatrix} 1.000 & -0.018 & -0.004 \\ -0.117 & 2.058 & 0.132 \\ -0.159 & 0.166 & 1.826 \end{bmatrix}, M_{QPCB} = \begin{bmatrix} 1.000 & 0.270 & 0.192 \\ -0.009 & 2.270 & 0.484 \\ -0.396 & 0.416 & 2.393 \end{bmatrix}.$$

$$(5.16)$$

---

[7]Note that these adjustments are not applied to the 'separate' models, for which a completely separate set of parameters is estimated for the best choice compared to the worst choice.

Table 5.5: Results from models for the best-worst dataset

models		MNL	RRM	QPCA	QPCA	QPCB	QPCB
	best-worst adjustment	Eq. 5.11	Eq. 5.11	Eq. 5.12	Eq. 5.13	Eq. 5.12	Eq. 5.14
Single	parameters	10	10	11	11	12	12
	Log-likelihood	-5,802.67	-5,803.97	-5,805.38	-5,802.48	-5,943.13	-5,772.78
	BIC	11,695	11,698	11,709	11,704	11,994	11,653
Separate	parameters	20	20	22	22	24	24
	Log-likelihood	-5,666.81	-5,667.24	-5,617.67	-5,617.67	-5,606.92	-5,606.92
	BIC	11,513	11,514	11,433	11,433	11,429	11,429
Quantum Rotation 1	parameters	13	13	13	13	14	14
	Log-likelihood	-5,752.51	-5,681.13	-5,681.86	-5,619.15	-5,692.32	-5,615.13
	BIC	11,622	11,479	11,480	11,355	11,510	11,356
	LL Gain	36.9%	89.8%	65.8%	99.2%	74.6%	95.1%
Quantum Rotation 2	parameters	15	15	16	16	17	17
	Log-likelihood	-5,731.11	-5,671.29	-5,672.43	-5,614.50	-5,662.90	-5,593.01
	BIC	11,597	11,477	11,488	11,372	11,478	11,338
	LL Gain	52.7%	97.0%	70.8%	101.7%	83.3%	108.4%
Quantum Rotation 3	parameters	16	16	19	19	20	20
	Log-likelihood	-5,730.17	-5,671.00	-5,672.43	-5,610.95	-5,662.29	-5,590.82
	BIC	11,604	11,485	11,515	11,392	11,504	11,361
	LL Gain	53.4%	97.3%	70.8%	103.6%	83.5%	109.7%

The impact of these matrices on the probability with which each alternative is chosen is demonstrated in Table 5.6. We start with a base scenario of two alternatives having a length of projection of 3 and one alternative having a length of 5, and consider how the probability of each alternative changes after a quantum rotation. For both models, a key change is that the probability of

Table 5.6: The impact on the probability of alternatives being chosen after quantum rotations

	Scenario 1	Scenario 2	Scenario 3
	Length of projection		
Alternative 1	5	3	3
Alternative 2	3	5	3
Alternative 3	3	3	5
	Initial Probability		
Alternative 1	58.1%	20.9%	20.9%
Alternative 2	20.9%	58.1%	20.9%
Alternative 3	20.9%	20.9%	58.1%
	New probability under QPCA		
Alternative 1	28.0%	5.6%	6.4%
Alternative 2	41.2%	71.6%	31.3%
Alternative 3	30.9%	22.8%	62.3%
	New probability under QPCB		
Alternative 1	27.2%	9.6%	9.0%
Alternative 2	45.1%	64.6%	33.6%
Alternative 3	27.7%	25.8%	57.4%

choosing alternative 1 decreases. As alternative 1 is the status quo alternative in this dataset, this suggests that an individual is more likely to choose the status quo as a best choice than a worst choice. Additionally, alternative 2 is more likely to be chosen than alternative 3 (as the worst). The higher values for elements $M_{2,3}$ and $M_{3,2}$ for M_{QPCB} compared to M_{QPCA} results in scenarios 2 and 3 being less deterministic after a rotation for QPCB compared to QPCA.

4.3.2 Contextual and ordering effects

Our next test of quantum rotation is to investigate its ability to capture contextual effects. For example, our second UK dataset has some choice sets with the cheaper alternative shown first, and some with the faster alternative

shown first. Additionally, cost is sometimes on the left and sometimes on the right. Whilst we could again use a full set of different parameters for the four different scenarios, quantum rotations could also be used. Two different quantum rotations are required, one for the order of the alternatives and the other for the order of the attributes. If a quantum rotation improves the model, it implies that the choice scenarios cannot be treated as equivalent if the context in which the choice is presented changes. We again test single and separate sets of parameters as well as a two versions of quantum rotations. The basic MNL and RRM models use parameters for cost and time as well as having one alternative specific constant[8]. QPCA and QPCB additionally have a second alternative specific constant[9] and QPCB additionally has two sensitivity parameters. The separate parameter models simply have four times as many parameters (a set of parameters for each combination of attribute and alternative order). The first set of rotations use diagonal matrices, resulting in one extra parameter per rotation for the quantum models and two extra per rotation for MNL and RRM. The second set uses matrix rotations with full sets of free parameters. The results of these models are given in Table 5.7.

Table 5.7: Results from applying quantum rotations to models for UK dataset 2

	models	MNL	RRM	QPCA	QPCB
Single	parameters	3		4	5
	Log-likelihood	-9,603.17		-9,369.61	-9,210.60
	BIC	19,235		18,778	18,469
Separate	parameters	12		16	20
	Log-likelihood	-9,584.18		-9,353.15	-9,189.43
	BIC	19,284		18,860	18,571
Quantum Rotation 1	parameters	7	7	6	7
	Log-likelihood	-9,593.83	-9,591.44	-9,359.81	-9,201.74
	BIC	19,255	19,250	18,777	18,471
	LL Gain	49.2%	61.8%	59.6%	41.9%
Quantum Rotation 2	parameters	11	11	10	11
	Log-likelihood	-9,592.17	-9,588.88	-9,355.14	-9,198.82
	BIC	19,290	19,284	18,806	18,503
	LL Gain	57.9%	75.3%	87.9%	55.7%

For all models, it appears that using separate sets of parameters results in an improvement in model fit. Whilst the quantum rotation models are not as successful as capturing the difference between the contextual situations,

---

[8] Note that whilst these models perform identically when there are only two alternatives, this is not the case when a quantum rotation is applied. This is a result of differing levels of impact on the difference in utility of the two alternatives.

[9] This does not cause identification issues as increasing both values here simply results in a less deterministic choice. As a contrast to the previous models in this chapter, the 2nd alternative specific constant results in a significant improvement for both quantum models for this dataset.

full matrix rotations for QPCA result in a model that is only 2 log-likelihood units worse than a model with a full set of separate parameters, which has 6 additional parameters. Consequently, the quantum rotation models achieve the best BIC values. Notably, both quantum models significantly outperform the MNL and RRM models. The parameter estimates for the quantum rotation matrices for changing from having the cheaper alternative on the left (first) to on the right (second) are:

$$M_{QPCA} = \left[\begin{array}{cc} 1.000 & 0.029 \\ -0.014 & 1.130 \end{array} \right], M_{QPCB} = \left[\begin{array}{cc} 1.000 & -0.053 \\ 0.102 & 1.225 \end{array} \right]. \quad (5.17)$$

The estimates for the quantum rotation matrices for changing from having the travel time first to having the travel cost first are:

$$M_{QPCA} = \left[\begin{array}{cc} 1.000 & 0.101 \\ -0.001 & 1.097 \end{array} \right], M_{QPCB} = \left[\begin{array}{cc} 1.000 & 0.043 \\ -0.008 & 0.898 \end{array} \right]. \quad (5.18)$$

We use similar scenarios before to test the impact of these rotations on the probabilities of picking the first or second alternatives. Using lengths of 3 and 5 for the base case (with the cheaper alternative first and the travel time first), the respective probabilities for choosing the alternatives are 27% for the alternative with length 3 and 73% for the alternative with length 5 (See Table 5.8). For all scenarios and all models, alternative 2 is always more likely to be chosen if it is the cheaper alternative, demonstrating that there is a bias towards picking the cheaper alternative. Whilst there is little change in the probabilities under QPCA for having the travel cost or travel time first, QPCB rotation parmeters results in alternative 1 being picked more often if travel cost is listed first. This is a result of having a value lower than 1 in matrix element $M_{QPCB_{2,2}}$.

4.3.3 Validation results for quantum rotations

We now again test for overfitting, this time for models for which full quantum rotations have been applied. We test both the quantum rotations from best to worst choice for our best-worst dataset and both rotations for the 2nd UK dataset. In all cases, the first 80% of the dataset is used for model estimation, with the remaining 20% used for validation. The results of these models are given in Table 5.9. Crucially, it appears that for both datasets, neither QPCA or QPCB overfits the data. It is in these models that we see more of a difference between the two quantum models, with QPCB outperforming QPCA in 7 out of 10 validation subsets.

4.4 Models for moral choice data

In our final results section, we apply variations of quantum models with and without quantum rotations to both moral choice datasets, demonstrating how

Table 5.8: Probabilities of picking alternatives after quantum rotations under both QPCA and QPCB

		QPCA		QPCB	
		Scenario 1	Scenario 2	Scenario 1	Scenario 2
		Length of projection			
Slow & cheap		3	5	3	5
Fast & expensive		5	3	5	3
		Cheaper alternative first, travel time first			
Slow & cheap		26.5%	73.5%	26.5%	73.5%
Fast & expensive		73.5%	26.5%	73.5%	26.5%
		Cheaper alternative second, travel time first			
Slow & cheap		23.9%	70.1%	15.3%	57.2%
Fast & expensive		76.1%	29.9%	84.7%	42.8%
		Cheaper alternative first, travel cost first			
Slow & cheap		29.0%	72.3%	34.1%	78.9%
Fast & expensive		71.0%	27.7%	65.9%	21.1%
		Cheaper alternative second, travel cost first			
Slow & cheap		26.7%	69.0%	21.5%	64.6%
Fast & expensive		73.3%	31.0%	78.5%	35.4%

Table 5.9: Results from holdout samples for the different models under a full quantum rotation for the best-worst and 2nd UK datasets

Best-worst dataset					
Model		MNL	RRM	QPCA	QPCB
Full Dataset	Estimate	-5,730	-5,671	-5,611	-5,591
Subset 1	Estimate	-4,563	-4,515	-4,457	-4,439
	Forecast	-1,172	-1,157	-1,157	-1,156
Subset 2	Estimate	-4,606	-4,573	-4,520	-4,505
	Forecast	-1,129	-1,103	-1,096	-1,091
Subset 3	Estimate	-4,534	-4,480	-4,430	-4,431
	Forecast	-1,203	-1,198	-1,186	-1,188
Subset 4	Estimate	-4,628	-4,578	-4,536	-4,519
	Forecast	-1,105	-1,095	-1,078	-1,074
Subset 5	Estimate	-4,570	-4,522	-4,485	-4,473
	Forecast	-1,165	-1,154	-1,131	-1,122
UK dataset 2					
Model		MNL	RRM	QPCA	QPCB
Full Dataset	Estimate	-9,592	-9,589	-9,355	-9,199
Subset 1	Estimate	-7,722	-7,721	-7,491	-7,386
	Forecast	-1,873	-1,871	-1,866	-1,816
Subset 2	Estimate	-7,667	-7,665	-7,475	-7,337
	Forecast	-1,928	-1,927	-1,882	-1,864
Subset 3	Estimate	-7,546	-7,544	-7,445	-7,374
	Forecast	-2,079	-2,077	-1,914	-1,827
Subset 4	Estimate	-7,671	-7,668	-7,484	-7,397
	Forecast	-1,924	-1,924	-1,873	-1,903
Subset 5	Estimate	-7,728	-7,724	-7,516	-7,451
	Forecast	-1,869	-1,870	-1,841	-1,845

they can capture changes in perspective in moral contexts.

4.4.1 A quantum rotation for taboo trade-offs

For the first set of models tested for the taboo trade-off dataset, we do not include any parameters to control for the presence of a taboo trade-off. We test a logit model as a comparison against two quantum models specified by Equations 5.10 and 5.3 respectively. Whilst models based on sine and cosine functions will likely not work for three or more alternatives (see Table 5.1 for an explanation of this), we only have two attribute levels in this dataset (which additionally makes the use of QPCB less desirable). The results of these models are given by Table 5.10. For both the quantum pairwise comparison model (QPCA) and the trigonometric quantum model (TQ), we find that the impact of the additional alternative specific constant is insignificant, thus we have 5 parameters for each of the models. In this case, neither basic quantum model performs as well as the basic logit model. Whilst QPCA appears to find nearly identical ratios for the estimates for the different attribute coefficients, the trigonometric quantum model assigns a higher importance to travel time.

Table 5.10: Results of basic models applied to the taboo trade-off dataset

Model	Pars.	LL	BIC	Ratios		
				Tax/Time	Injuries/Time	Deaths/Time
Logit	5	-721.23	1,479.30	1.88	2.14	1.52
QPCA	5	-725.40	1,487.64	1.90	2.13	1.49
TQ	5	-722.94	1,482.72	1.68	1.85	1.32

Next, we apply quantum rotations. In this case, a quantum rotation is used to capture the shift in perception of the alternatives in the presence of a taboo trade-off. Thus, if the decision-maker can decrease travel time or tax at the cost of increasing the number of fatalities or serious injuries, a quantum rotation is applied to the estimated projection lengths. As multiplying each projection by the same factor will result in no change in the probabilities with which the alternatives are chosen, the first element of the rotation matrix must be fixed to one. We then test three different rotations with varying amounts of flexibility: a diagonal matrix (QR1), a symmetrical matrix (QR2) and a fully flexible matrix (QR3). The results of these models are given in Table 5.11, with the rotation matrix elements that are estimated given in bold.

For both types of quantum models, we see a significant improvement in log-likelihood from the addition of 1 rotation parameter. However, whilst the addition of further parameters continues to improves log-likelihood, this results in a worse BIC. Crucially, Chorus et al. (2018)'s 'Generic Taboo Trade-

Table 5.11: Results of the quantum rotation models applied to the taboo trade-off dataset

Quantum Pairwise Comparison A				Quantum rotation matrix			
Version	Parameters	Log-likelihood	BIC	M[1,1]	M[1,2]	M[2,1]	M[2,2]
Basic	5	-725.40	1,487.64	1.00	0.00	0.00	1.00
QR1	6	-718.78	1,481.77	1.00	0.00	0.00	**1.40**
QR2	7	-716.90	1,485.37	1.00	0.18	**0.18**	1.18
QR3	8	-716.49	1,491.92	1.00	**0.32**	**-0.06**	1.55
Trigonometric Quantum				Quantum rotation matrix			
Version	Parameters	Log-likelihood	BIC	M[1,1]	M[1,2]	M[2,1]	M[2,2]
Basic	5	-722.94	1,482.72	1.00	0.00	0.00	1.00
QR1	6	-717.33	1,478.86	1.00	0.00	0.00	**1.29**
QR2	7	-714.21	1,479.99	1.00	0.12	**0.12**	1.32
QR3	8	-713.63	1,486.20	1.00	**0.30**	**-0.17**	1.84

Off Aversion' (TTOA) model (which is a logit model with an additional parameter that adds a penalty for the presence of a taboo trade-off) achieves a log-likelihood of -719.47 and a BIC of 1,483.15, values which are both worse than the QR1 models tested here. As a contrast to the earlier results in this chapter, the trigonometric quantum model outperforms the quantum pairwise comparison model. This is in part possibly due to the fact that there is less variation in the attribute levels in the choice tasks. Notably, all rotation models have a value greater than 1 for element $M_{2,2}$, which, all else being equal, would mean that the decision-maker is more likely to choose the 2nd alternative (which is the status quo alternative) in the presence of a taboo trade-off (in line with the results of Chorus et al. 2018). However, the different estimates for $M_{1,2}$ and $M_{2,1}$ mean that the overall impact is more complex for QR2 and QR3. Table 5.12 gives the probability of supporting the new policy before the quantum rotation and after it for both QR3 models. Crucially, in both models, the predicted shares of supporting a policy are closer on average to the observed shares than in the TTOA model.

For QPCA, the probabilities tend to become less extreme, whereas they often become more extreme under the TQ model. Crucially, however for both models, if the lengths are similar, then the probability for the status quo increaseses in the presence of a taboo trade-off. Additionally, both models have smaller mean absolute deviations from the true share of support than the TTOA model.

4.4.2 A quantum rotation when you consider your partner

For the datasets on willingness to accept longer commutes for better salaries, there are two distinct choice sets: the first only includes factors impacting the decision-maker, the second includes impacts on the partner. Consequently, we first test the difference between models that treat the two sets as the same

Table 5.12: Impact on probabilities as a result of quantum rotations

Scenario	Attributes				Taboo Trade-Off?	Share of support					
	Tax	Time	Injuries	Deaths		Observed	TTOA	QPCA QR3 before	after	TQ QR3 before	after
1	-	-	-	-	No	98.0%	92.0%	96.9%		97.4%	
2	-	-	-	+	Yes	68.7%	71.0%	74.9%	66.4%	77.5%	67.0%
3	-	-	+	+	Yes	29.3%	21.0%	19.2%	21.8%	31.7%	24.6%
4	-	+	+	+	Yes	11.1%	9.0%	5.1%	11.3%	11.3%	12.1%
5	+	+	+	+	No	2.0%	1.0%	0.6%		2.1%	
6	+	+	-	-	No	62.6%	63.0%	64.9%		63.4%	
7	+	+	-	-	No	44.4%	37.0%	37.7%		38.1%	
8	+	+	+	-	No	4.0%	6.0%	2.3%		3.4%	
9	-	+	+	+	Yes	42.4%	46.0%	48.8%	42.7%	53.6%	41.0%
10	+	-	+	+	Yes	15.2%	15.0%	12.2%	16.8%	18.3%	16.3%
11	-	-	+	+	Yes	46.5%	56.0%	58.6%	50.7%	64.4%	51.3%
12	-	+	-	-	No	80.8%	81.0%	85.6%		83.7%	
13	-	+	+	-	Yes	30.3%	31.0%	31.5%	30.2%	39.1%	29.7%
14	+	-	-	+	Yes	22.2%	26.0%	24.2%	25.2%	30.7%	24.0%
15	+	-	+	+	Yes	5.1%	4.0%	0.4%	5.6%	1.2%	4.8%
16	+	+	-	+	No	7.1%	11.0%	7.5%		10.6%	
Mean absolute deviation from true share of support (percentages; all choice tasks)							3.03	2.41	2.21	2.01	1.93
Mean absolute deviation from true share of support (percentages; taboo tasks only)							2.68	2.21		1.93	

(and thus have the same parameter estimates for the alternative specific constants and utility coefficients for the impact of changes in salary and travel time for the decision-maker, across both choice sets). The second set of models is essentially made up of two separate components, as it has entirely different sets of parameters for the different choice sets. We test both random regret minimisation models (which were previously demonstrated (Hess et al., 2014) to be effective for this dataset due to the presence of an indifferent option) as well as the quantum pairwise comparison model based on Equation 5.10. All models simply estimate a constant for the regret (for RRM) or projection length (for the quantum model) for the indifferent alternative. The results of these models are given in Table 5.13.

Table 5.13: Models with and without separate sets of parameters for the two different choice sets

Model	Separate Pars.	Free Parameters	Log-likelihood	BIC
RRM	No	6	-12,784.21	25,628
RRM	Yes	10	-12,426.71	24,952
QPCA	No	7	-12,624.13	25,318
QPCA	Yes	12	-12,289.38	24,698

Regardless of whether RRM and QPCA are compared with or without separate parameters, the results indicate that the quantum models provide a large gain in model fit as well as substantially lower BIC values. Additionally, both QPCA and RRM find clear evidence that separate sets of parameters can be used to improve model fit, demonstrating that there is an inconsistency in how a decision-maker considers factors impacting themselves compared to when there are also factors impacting their partner.

Crucially, however, this inconsistency or 'change of mindset' incurred through changing from thinking about just yourself compared to yourself *and your partner* could be captured by a quantum rotation. Thus, for our quantum rotation models, we estimate a single set of coefficients that apply to choices made in both choice sets, and instead include a quantum rotation matrix for adjusting the projection lengths appropriately when additionally considering travel time and salary changes for the partner. We again test diagonal, symmetric and fully flexible matrices. The results of these models are given in Table 5.14.

Whilst additional flexibility for the quantum rotation matrices did not significantly improve model fit for the taboo trade-off dataset, we see a consistent improvement both in log-likelihood and BIC for the quantum rotation models here. Additionally, in line with previous results in this chapter, quantum rotation models can again provide better model fit than a model that allows for a separate set of parameter values (although in this case the BIC

Table 5.14: Results from quantum rotations when considering travel time and salary changes for your partner

#	Model	Parameters	Log-likelihood	BIC	Log-likelihood Gain	
					Over basic model	As percentage of Model 2
1	Basic	7	-12,624.13	25,318	-	-
2	Separate	12	-12,289.38	24,698	334.75	100%
3	QR1	9	-12,436.03	24,961	188.10	56%
4	QR2	12	-12,332.37	24,784	291.76	87%
5	QR3	15	-12,278.06	24,705	346.07	103%

is slightly better for a model with separate parameters). This suggests that these models provide a suitable framework for capturing changes in choice context.

For the most complex of these models, the rotation matrix estimated is:

$$M = \begin{bmatrix} 1.000 & -0.464 & 4.834 \\ -0.046 & 0.116 & 2.631 \\ 0.151 & -0.068 & -0.651 \end{bmatrix}, \tag{5.19}$$

where the first alternative is to stay with the status quo, the second is to increase travel times and salaries and the third is the indifferent option. The high values for $M_{1,3}$ and $M_{2,3}$ indicate that if an individual is indifferent for a choice scenario involving just changes for themself, then they will likely not still be indifferent if there are additionally changes for their partner.

5 Conclusions

In this chapter, we move away from the tried and tested alternatives to random utility maximisation by considering ideas first developed in quantum physics. With the probability framework developed in quantum physics having made a successful transition to cognitive psychology, we look at whether it can be operationalised into a choice model framework. Under quantum logic, a decision-maker has some 'indefinite state' regarding their preferences over alternatives, from which the probabilities of each alternative can be inferred. Thus a key component of this chapter is the development of specifications for the indefinite state.

We find two new possible suitable specifications for the 'indefinite state' which allow us to incorporate quantum logic within a choice model. The first uses an adapted specification based on random regret minimisation. The second uses a variation of the specification of the mean drift rates within a multi-attribute linear ballistic accumulator model. We find that our quantum pairwise comparison version A (QPCA) model provides good model fit and outperforms multinomial logit (MNL) and random regret minimisation (RRM) across three simple route choice datasets as well as providing good out-of-sample fit for the most complex of these datasets. Additionally, the quantum pairwise comparison version B (QPCB) model also provides good fit for all three datasets. Whilst QPCB has better model fit, QPCA provides more robust parameter estimates. The third quantum framework based on sines and cosines, appears unsuitable, as there are basic probability issues for three or more alternatives. Nevertheless, the positive results from our initial tests on the QPCA and QPCB models suggest that there is ample scope for models with a quantum framework to be used within travel behaviour modelling. They are simple to run and estimate, meaning that they could

be applied to a wide range of choice scenarios. However, for these models to make a transition into large-scale modelling, an alternative specification would need to be defined to avoid the same pitfall of random regret minimisation for large numbers of alternatives: a comparison between every pair of alternatives quickly becomes computationally infeasible.

Another issue with the current specifications of the models is that it could be argued that there is no real 'quantum insight' in the model structure. QPCA is in essence a random regret minimisation model with a different error structure. However, results from our quantum rotation models suggest that there are many benefits of bringing quantum logic into choice models. Most crucially, it appears that they can accurately capture a change in decision context, both within traditional route choice datasets and also when a moral element enters the dimension of choice.

Crucially, our best performing models for the best-worst dataset and the contextual choice dataset, after allowing for model complexity to be taken into account, are the rotated QPCA and rotated QPCB models. This suggests that there is some merit in the concept of quantum rotation, which suggests a different set of basis vectors are required for different choice situations. For example, the rotation works well for capturing the difference between best and worst choice. Despite the fact that the best-worst choices are related, quantum rotation suggests that these choices are in fact incompatible: the choices cannot be made at the same time and consequently they may not follow the classical probability law of distributivity. This means that different choices may be observed depending on whether the decision-maker chooses the best or worst alternative first. Similar results are also obtained for the moral choice datasets. Whilst these results are positive, it is not clear that they are distinctly *better* than those of the route choice datasets. This implies that we cannot necessarily attribute the success of the quantum models for the moral data to the fact that there are moral components. This is particularly clear from the result that our quantum model already has better model fit than the random regret model for the 2nd moral choice dataset tested in this chapter *before* the moral component was captured through a quantum rotation. Further tests of quantum logic based models could further enlighten whether they are models that are particularly suited to moral decision-making, or whether they are suitable for decision-making in general.

Additionally, as moral decision making is becoming an interesting area of application for choice modelling, there is a need for the development of appropriate model specifications. Further models could consider different sorts of moral choice data. For example, quantum models may be well suited for modelling choices made in 'moral machine' choice tasks. Results may also differ significantly for revealed preference datasets, with the concern of external validity of stated moral choices (Bauman et al., 2014) still an issue that has yet to have been addressed.

Overall, the results in this chapter demonstrate that there is a large amount of scope for future work within choice modelling. For example, large-scale models frequently aim to understand a series of related, sequential choices. Given the ability of quantum rotation to capture the translation between best and worst choices, it theoretically should also work for a larger sequence of related choices where continuously adding on separate sets of parameters may not be possible. Thus ordering effects and state dependence may be well captured by models within a quantum framework. Furthermore, it may be possible to mitigate the impacts of contextual effects by applying the appropriate quantum rotation from other models that account for the same effect. Future efforts could also compare quantum frameworks against other models that are specifically designed to deal with contextual effects, such as prospect theory. These future possibilities combined with the positive results across both route and moral choice datasets mean that this chapter serves as a proof-of-concept that choice models with a quantum logic framework have vast potential, both within route and moral choice scenarios and more generally.

Acknowledgements

References

Avineri, E., Bovy, P.H., 2008. Identification of parameters for a prospect theory model for travel choice analysis. Transportation Research Record 2082, 141–147.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., Rahwan, I., 2018. The moral machine experiment. Nature 563, 59.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J., Bierlaire, M., 2008. Income and distance elasticities of values of travel time savings: New swiss results. Transport Policy 15, 173–185.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M.O., Chintakayala, P.K., Choudhury, C., Daly, A., Dekker, T., et al., 2017. New appraisal values of travel time saving and reliability in great britain. Transportation , 1–39.

Bauman, C.W., McGraw, A.P., Bartels, D.M., Warren, C., 2014. Revisiting external validity: Concerns about trolley problems and other sacrificial

dilemmas in moral psychology. Social and Personality Psychology Compass 8, 536–554.

Beck, M.J., Hess, S., 2016. Willingness to accept longer commutes for better salaries: Understanding the differences within and between couples. Transportation Research Part A: Policy and Practice 91, 1–16.

Ben-Akiva, M., McFadden, D., Train, K., et al., 2019. Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. Foundations and Trends® in Econometrics 10, 1–144.

Bierlaire, M., Thémans, M., Zufferey, N., 2010. A heuristic for nonlinear global optimization. INFORMS Journal on Computing 22, 59–70.

Birkhoff, G., Von Neumann, J., 1936. The logic of quantum mechanics. Annals of mathematics , 823–843.

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

Bruza, P.D., Wang, Z., Busemeyer, J.R., 2015. Quantum cognition: a new theoretical approach to psychology. Trends in cognitive sciences 19, 383–393.

Busemeyer, J.R., Pothos, E.M., Franco, R., Trueblood, J.S., 2011. A quantum theoretical explanation for probability judgment errors. Psychological review 118, 193.

Busemeyer, J.R., Townsend, J.T., 1992. Fundamental derivations from decision field theory. Mathematical Social Sciences 23, 255–282.

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

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

Chorus, C.G., 2015. Models of moral decision making: Literature review and research agenda for discrete choice analysis. Journal of choice modelling 16, 69–85.

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

References

Chorus, C.G., Pudāne, B., Mouter, N., Campbell, D., 2018. Taboo trade-off aversion: A discrete choice model and empirical analysis. Journal of choice modelling 27, 37–49.

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

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

Giergiczny, M., Dekker, T., Hess, S., Chintakayala, P.K., 2017. Testing the stability of utility parameters in repeated best, repeated best-worst and one-off best-worst studies. European Journal of Transport and Infrastructure Research 17.

Guevara, C.A., Fukushi, M., 2016. Modeling the decoy effect with context-rum models: Diagrammatic analysis and empirical evidence from route choice sp and mode choice rp case studies. Transportation Research Part B: Methodological 93, 318–337.

Heisenberg, W., 1927. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. Zeitschrift für Physik 43, 172–198. doi:10.1007/BF01397280.

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

Hess, S., Beck, M.J., Chorus, C.G., 2014. Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives. Transportation Research Part A: Policy and Practice 66, 1–12.

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

Hess, S., Daly, A., Dekker, T., Cabral, M.O., Batley, R., 2017. A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. Transportation Research Part B: Methodological 96, 126–149.

Hess, S., 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., 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, 847–863.

References

Hess, S., 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.

Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39, 565–591.

Hughes, R.I., 1992. The structure and interpretation of quantum mechanics. Harvard university press.

Kennard, E.H., 1927. Zur quantenmechanik einfacher bewegungstypen. Zeitschrift für Physik 44, 326–352.

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

Leong, W., Hensher, D.A., 2014. Relative advantage maximisation as a model of context dependence for binary choice data. Journal of choice modelling 11, 30–42.

Lipovetsky, S., 2018. Quantum paradigm of probability amplitude and complex utility in entangled discrete choice modeling. Journal of choice modelling 27, 62–73.

Mahieu, P.A., Crastes, R., Louviere, J., Zawojska, E., et al., 2016. Rewarding truthful-telling in stated preference studies. Technical Report.

Masiero, L., 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, 349–358.

Planck, M., 1901. Ueber das gesetz der energieverteilung im normalspectrum. Annalen der Physik 309, 553–563. doi:10.1002/andp.19013090310, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19013090310.

Pothos, E.M., Busemeyer, J.R., 2009. A quantum probability explanation for violations of 'rational' decision theory. Proceedings of the Royal Society B: Biological Sciences 276, 2171–2178.

Pothos, E.M., Busemeyer, J.R., 2013. Quantum principles in psychology: the debate, the evidence, and the future. Behavioral and Brain Sciences 36, 310–327.

Cunha-e Sá, M.A., Madureira, L., Nunes, L.C., Otrachshenko, V., 2012. Protesting and justifying: a latent class model for contingent valuation with attitudinal data. Environmental and Resource Economics 52, 531–548.

Savage, L.J., 1954. The foundations of statistics; jon wiley and sons. Inc.: New York, NY, USA .

Seetharaman, P., 2003. Probabilistic versus random-utility models of state dependence: an empirical comparison. International Journal of Research in Marketing 20, 87–96.

Stathopoulos, A., 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, 1673–1689.

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

Trueblood, J.S., Busemeyer, J.R., 2011. A quantum probability account of order effects in inference. Cognitive science 35, 1518–1552.

Trueblood, J.S., Busemeyer, J.R., 2012. Quantum information processing theory, in: Encyclopedia of the Sciences of Learning. Springer, pp. 2748–2751.

Trueblood, J.S., Pothos, E.M., Busemeyer, J.R., 2014b. Quantum probability theory as a common framework for reasoning and similarity. Frontiers in psychology 5, 322.

Turner, B.M., Schley, D.R., Muller, C., Tsetsos, K., 2018. Competing theories of multialternative, multiattribute preferential choice. Psychological Review 125, 329.

Tversky, A., 1977. Features of similarity. Psychological Review 84, 327.

Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological review 90, 293.

Vitetta, A., 2016. A quantum utility model for route choice in transport systems. Travel Behaviour and Society 3, 29–37.

Wang, Z., Busemeyer, J.R., Atmanspacher, H., Pothos, E.M., 2013. The potential of using quantum theory to build models of cognition. Topics in Cognitive Science 5, 672–688.

Yu, J.G., Jayakrishnan, R., 2018. A quantum cognition model for bridging stated and revealed preference. Transportation Research Part B: Methodological 118, 263–280.

Chapter 6

New models for dynamic choice contexts: further steps towards bridging choice modelling with mathematical psychology

Thomas Hancock[1], Stephane Hess[1], & Charisma Choudhury[1]

Abstract

Modelling decisions in dynamic choice contexts, where the decision maker has to make a single or a series of choices in a changing scenario (e.g. driving manoeuvres, route choice, investment decisions, etc.) is much more challenging than in static scenarios. These decisions have typically been based on models that split time into discrete intervals, estimate utilities for each time point, then use lagged variables to account for the dynamic nature of the scenario. This is despite the simultaneous development of 'sequential sampling' (accumulator) models, that attempt to capture the accumulation of preference over time. The application of such models have however been limited to scenarios with static attributes. Decision field theory (DFT), one such model which assumes that the preferences for different alternatives update stochastically over time, has recently made the transition into mainstream choice modelling. Whilst this model has been effective at modelling, for example, route choice, it has not yet been tested on choice scenarios where the attributes of alternatives update over time. In this paper, we discuss the relative benefits of using an accumulator model for dynamic choice contexts, as well as developing and operationalising both DFT and models based on quantum probability frameworks for dynamically changing attributes in the context of driving behaviour problems. We demonstrate that analytical and simulated

---

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)

versions of DFT as well as quantum models provide a better account of behaviour than models based on more traditional methods. Whilst the models developed in this paper are preliminary in nature, we demonstrate that there is immense potential in the development of such models for dynamic choice scenarios.

1   Introduction

The vast majority of applications in choice modelling look at individual choices in isolation, where a decision maker chooses amongst a finite set of alternatives from a static choice set, i.e. one where the alternatives and their attributes do not change over time. A relatively small departure from that context comes in work looking at the relationship between individual choices over time, largely modelled through state dependence (Bjørner and Leth-Petersen, 2005; Kitamura and Bunch, 1990).

A rather different set of circumstances applies in the context where a choice is made in a scenario where the composition of the choice set and/or the attributes describing the alternatives changes during the time interval when the choice takes place. Work with dynamic discrete choice models (DDCM) looks at situations where the set of alternatives or their attribute levels can change (Cirillo and Xu, 2011). Dynamic problems are often formulated as Markov decision processes (Rust, 1994) under which the next system state is dependent on a Markov transition probability. This has typically been applied to consumer problems such as the purchase of digital camcorders (Rysman et al., 2009) or the purchase of new automobiles (Schiraldi, 2011). To a large extent, much of this work is based on the idea of splitting the choice process into a number of sequential choices. This is however quite different from the situation where only a single choice is made, but where the characteristics of the choice set evolve during the process of the decision maker making that choice. In addition, choice contexts in some cases can change much more rapidly than for the longer term scenarios discussed above. This is especially the case in transport. In particular, the decision to merge into a different lane is dependent on the size of the gap in the target lane, which changes rapidly as the vehicles in front and behind change speed. Previous efforts at modelling merging decisions and gap acceptance problems have also included logit models (Marczak et al., 2013; Polus et al., 2005; Rossi et al., 2013), but are largely static in nature. A different application of dynamic models comes in route choice modelling, with Fosgerau et al. (2013) demonstrating that a dynamic model for sequential link choices can be used effectively to capture route choice in a network and Mai et al. (2015) further demonstrating that correlation across error terms can be included for such problems. This work however also just replaces a single choice of an

overall route by a recursive set of individual decisions, where in addition, the
attributes of the network are stable over time.

In DDCMs, the continuous time horizon is typically split into separate in-
dividual choice situations, where the decision to take a given action (possibly
out of a set of actions) or not, is modelled at each point in time (Aguirre-
gabiria and Mira, 2010). In contrast to these 'static' models, many 'dynamic'
choice models have been developed in mathematical psychology, where the
preference of an alternative updates over time as the decision-maker delib-
erates on the set of alternatives (Bhatia, 2013; Busemeyer and Townsend,
1992; Ratcliff and Rouder, 2000; Usher and McClelland, 2001). In these 'ac-
cumulator' models, however, the dynamic process can thus be estimated as a
single continuous process, where the preferences change over time and where
an action is taken when a certain preference threshold is reached.

These 'process' or 'sequential sampling' models aim to better understand
the choice by attempting to additionally consider the decision-maker's de-
liberation process that occurs whilst they consider their alternatives. This
results in models that can accurately predict response times (Brown and
Heathcote, 2008; Diederich and Busemeyer, 2003) as well as choice outcomes.
Typically, a key focus of accumulator models has been on how well they pre-
dict various contextual effects (Noguchi and Stewart, 2018; Trueblood et al.,
2014; Tsetsos et al., 2010), with the transition of such models into 'main-
stream' choice modelling limited due to computational complexities of such
models (Otter et al., 2008). However, recent work has demonstrated that
decision field theory (DFT), a stochastic, dynamic model for understanding
choices, and the multiattribute linear ballistic accumulator model (MLBA),
a model where the preference for an alternative 'drifts' linearly, can both be
used effectively to model travel and consumer choice problems (see Chapter
4).

The focus in the application of these models has however been on the dy-
namic updating of preferences (i.e. internal accumulation of evidence) rather
than looking at dynamically changing attributes of the alternatives/choice
sets. Whilst these models should theoretically be able to easily incorporate
updating of attributes, applications thus far have been limited to basic sce-
narios such as changes in dot direction in a dot perception task (Holmes et al.,
2016). While the use of DDCMs may be appropriate for longer term choices,
where a decision maker over time returns to a specific choice context and
thinks about it again, it seems in our view a less natural solution for situa-
tions where a decision maker focusses on a specific choice for a relatively short
period of time but where the attributes of that choice are not static. Numer-
ous examples exist. A driver aiming to merge onto a motorway may have
to assess a number of gaps, where the attributes of the gap (e.g. gap-size)
are changing over time, before finding a suitable gap to merge into. Alter-
natively, a traveller may be tracking hotel prices over multiple days, where

the prices are changing dynamically, before making a booking. Outside of transport, stock market investment decisions are evaluated over time, while share prices change continuously. This paper aims to bring together dynamic datasets and dynamic models, testing whether models and ideas developed in mathematical psychology can be used to better model decision probabilities in dynamic environments, where we particularly focus on decision field theory (DFT).

Under DFT, a decision-maker is assumed to consider different attributes for the different alternatives over the course of the deliberation process. In the standard implementation of DFT, this *sampling* takes information from a static set, i.e. the values for the different attributes and alternatives are kept constant during the deliberation. This is appropriate for example in the context where a decision maker looks at a shelf of products in a supermarket and keeps alternating between price and some quality attribute for comparing the alternatives, where these attributes remain static. It is however less appropriate for truly dynamic setting as it in essence assumes that a decision maker takes a snapshot of the information describing the alternatives at the outset, and then keeps comparing the alternatives internally using that information. It seems much more likely that there is in addition some *external* updating of the information, notwithstanding the possibility that the *internal* deliberation may operate at a higher frequency (i.e. thinking repeatedly about a specific gap size and speed differential before updating the information on this after it changes).

A simple example where a decision-maker considers the choice between two alternative routes helps illustrate the concept. The first alternative has a cost of £5 and travel time of 30 minutes and the second costs £10 and takes 20 minutes. This might result in the preferences for alternatives updating over time in line with the values given in Figure 6.1. In each timestep, one attribute is considered, and the alternative which performs better for that attribute gains in preference value. For example, in the 1st, 3rd and 4th deliberation (preference updating) timesteps, the decision-maker considers cost, and consequently the cheaper alternative quickly becomes favourable, even though the faster alternative was originally preferred. If the first alternative to reach a preference value of 5 is chosen, the decision-maker would then choose the cheaper alternative. If, however, the decision-maker did not come to a conclusion, they might reach a point where the attributes change. The example in Figure 6.1 demonstrates what might happen if tickets were reduced in price for the faster alternative (after 5 deliberation timesteps). This results in the faster alternative quickly becoming the preferred option.

Thus far, the only previous attempt to fit such a model to dynamic data (as far as the authors are aware) was made by Holmes et al. (2016). They used an adapted version of the linear ballistic accumulator model (where the preferences for alternatives grows linearly until the preference for one alter-

179
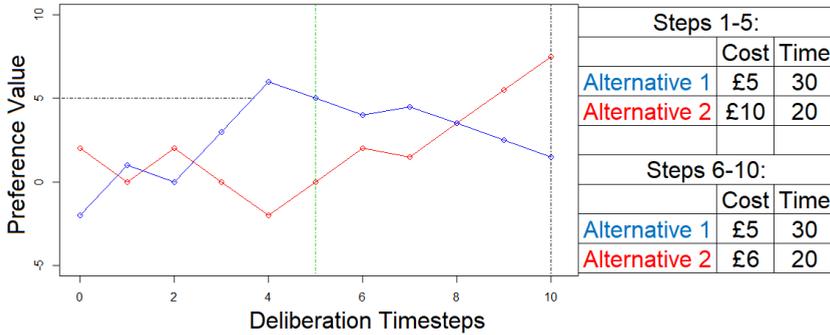
Fig. 6.1: An example deliberation process when attribute values change, under a DFT model

native reaches a threshold) to model decisions about moving dot perception tasks. However, many real-world dynamic choices are more complex, thus ready-made accumulator models with analytical solutions for calculating the probability of alternatives being chosen may not be possible. The aim of this paper is to introduce an effective dynamic choice model to predict choices under dynamic environments. We consider the use of decision field theory for understanding a driver's decision as to whether or not to merge onto a motorway. As well as testing a version with attributes staying constant for a set number of deliberation timesteps before updating (as depicted in Figure 6.1 and in line with the earlier point about internal deliberation operating at a higher frequency than external updating), we also compare a simulated approach where the attributes change with every deliberation timestep. We also compare these to more traditional 'static' models that use exponential and hyperbolic discounting functions to account for 'remembered' preferences evaluated at previous attribute values.

An alternative approach is to add a dynamic element to quantum probability models. These models are based on the operationalisation of quantum logic to create a different probability system to generate the probabilities with which choices (or actions) are made. Models with such a structure have recently made a significant impact in cognitive psychology (c.f. Busemeyer and Bruza 2012 for a review) and consequently have been demonstrated to show significant promise in choice modelling (in Chapter 5). The key assumption under quantum logic is that some pairs of decisions are 'incompatible', meaning that the probabilities of choosing different alternatives or actions in one decision are impacted by the choice made in the other. This results in the adoption of quantum logic allowing for convenient and elegant solutions to, for example, ordering effects (Trueblood and Busemeyer, 2011; Wang et al.,

2014). In this chapter, we consider a model based on Lipovetsky (2018)'s quantum choice model. In Chapter 5, we demonstrate that quantum rotations can be used to capture changes in choice contexts. In this chapter, we explore whether these rotations can also be used to capture changing choice contexts. Thus a natural extension is to test whether these models can also capture dynamically changing choice contexts.

The remainder of this paper is organised as follows. First, we give an outline of what a dynamic accumulator model for dynamic choice contexts could theoretically include. Next, we detail mathematically how DFT and quantum models could account for changing choice contexts. We then describe the US101 merging dataset and detail how our models are adapted for modelling the choice to merge into a new lane or not. We then report the results of applying different variations of the models to the data. We finish by giving some conclusions as well as detailing future steps for improving dynamic models.

2 Dynamic models: theory and elements to include

In this section, we first consider a theoretical example of preferences for alternatives changing over time whilst attributes of the alternatives also change. We then detail a number of features that a dynamic model could consider, before detailing how some of these can be implemented under a decision field theory model. Finally, we demonstrate how a model based on quantum probability can capture dynamic elements through quantum rotations.

2.1 A theoretical example

Whilst one method for visualising the preference for alternatives changing over time is demonstrated by Figure 6.1, another possibility is to consider differences between preferences, with many models already framing alternatives in this way (Busemeyer and Townsend, 1992; Krajbich et al., 2012; Noguchi and Stewart, 2018) such that thresholds are based on relative differences between the preferences for alternatives. For example, it is simple to reformulate our original example in terms of differences between the preferences, where this is shown in Figure 6.2, using the difference between the preferences for alternative 2 and alternative 1.

This example additionally considers the possibility of updating threshold values and the impact of dampening effects when attributes change. These are just two of a number of factors that a dynamic model could theoretically take into account, and we now look at these in turn:
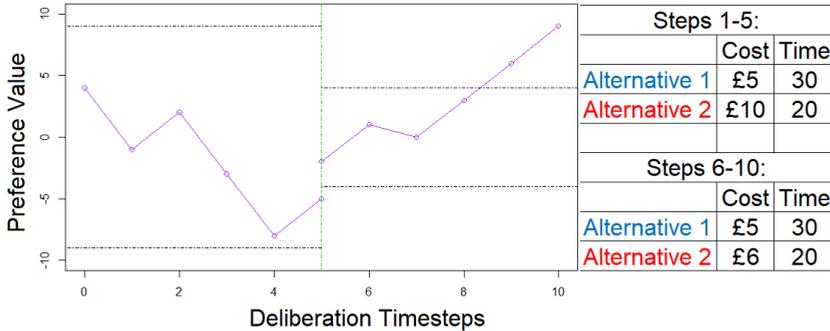
Fig. 6.2: DFT example with difference in preference values in a dynamic setting

Impact of deliberation time: For the model to be dynamic, it needs to have some element that can be related to real decision time. If it includes some number of 'preference updating steps' (as depicted in Figure 6.2) then the decision-maker's response time taken to make a choice can be incorporated by setting the number of deliberation steps as some function of this time. This has been considered for decision field theory for static attributes in Chapter 3.

Different rates for attribute updates compared to preference updates In the above example, there are 5 deliberation steps for each of the two different sets of attribute values. There could clearly be a large number of deliberation steps for a slowly changing scenario such as the one above, and this is particularly the case when the acquisition of new information requires some action or comes at some cost, with many models already incorporating information search costs (Drugowitsch et al., 2012; González-Valdés and de Dios Ortúzar, 2018; Kim et al., 2016). The number of deliberation steps used for each new set of attribute values may be much lower in a continuously changing scenario, and the internal processing speed may in some cases go as low as the external updating. The rate at which an individual samples the real world information may also change over time, with for example more rapid updating the closer the person comes to making a decision. Finally attributes may not be updated from visual stimuli with every preference updating step but could be interpolated towards an individual's expectation of how the attribute will change.

Processing speeds: In a rapidly changing context, the processing of attribute changes will also be a function of how quickly an individual can

process new information. For example, if there is suddenly a change in price for an ebay item very close to the ending time of the auction, a consumer may not react in time to increase their bid. For this reason, choices made over a short timeframe (such as driving decisions, for example) may be subject to both mental processing speeds and physical reaction times, with Holmes et al. (2016) finding a strong delay effect before a decision-maker integrated new information in a moving dot perception task.

Significant threshold changes: In Figure 6.2, the upper boundary represents the threshold for alternative 2 being chosen and the lower boundary represents the threshold for alternative 1 being chosen. Theoretically, when the attributes change, the thresholds may also change. Drift diffusion models have often considered threshold changes or 'collapsing boundaries' (Zhang et al., 2014), although recent evidence suggests that these models do not have an empirical advantage over models with a fixed threshold (Hawkins et al., 2015; Voskuilen et al., 2016).

Gradual threshold changes: Thresholds may also change more gradually than depicted in Figure 6.2. For example, a careful driver's preference may initially need to reach a high threshold for merging lanes if they only consider merging when the gap is particularly large. However, as the driver approaches the end of the slip road, their threshold may gradually decrease if they otherwise risk not merging into the correct lane, with drivers more likely to accept a gap if they are under stress (Paschalidis et al., 2018).

Dampening effects: If attributes change gradually, the impact of dampening may be small. However, a more significant change in attribute values such as shown in the example in Figure 6.2 may result in preference values resetting to zero or to their initial starting values. For example, an individual considering purchasing an item on ebay may reconsider an alternative that they had previously ruled out if the price is significantly reduced. Many process models already include 'decay' parameters (Roe et al., 2001; Usher and McClelland, 2001), but separate parameters may be required for gradual changes compared to sudden changes.

Changes in the set of available alternatives: While we have focussed on changes in attribute values, it is also possible that whilst the decision-maker is deliberating, one of the alternatives becomes unavailable. This may result in preferences resetting, or a large shift towards the most similar alternative. A typical transport example could be a cancelled train service.

Initial preferences towards alternatives: It is easily possible that there may be initial biases or preferences towards an alternative. For example, commuters will likely choose the same route to work that they chose the day before. With random utility models adept at capturing such effects and decision field theory also able to account for initial preferences (see Chapter 2), a dynamic model that accounts for changing attributes should also be able to account for initial preferences. This is particularly important given that more complex choice scenarios result in more frequent choice of the status quo (Boxall et al., 2009).

In the following section, we look at a steps towards including some of these factors in models based on decision field theory.

2.2 Decision field theory

One model that can already account for several of the above factors is decision field theory (DFT), which was first developed by Busemeyer and Townsend (1992, 1993). It is a stochastic, dynamic model for understanding choices between multiple alternatives. Under DFT, each alternative has a 'preference value' that stochastically updates over time as the decision-maker considers the different attributes for the alternatives. The decision-maker considers their alternatives for some number of deliberation timesteps until one of the alternatives reaches some internal threshold value or until the decision-maker runs out of time upon reaching some external threshold. These thresholds are represented graphically in Figure 6.1 by the horizontal dashed line (preference value = 5) and the vertical dashed line (deliberation timesteps = 10) respectively.

Many variations of DFT exist, with structures for models with internal thresholds (Busemeyer and Townsend, 1993) rather different to versions for multiple alternatives and multiple attributes, which have external thresholds (Berkowitsch et al., 2014; Roe et al., 2001). For the work in this paper, we consider models based on DFT models with external thresholds as this version offers a clearer method for the implementation of updating attribute values. Although analytical solutions for internal DFT are available (albeit only for choice scenarios with two alternatives, see Busemeyer and Townsend 1993), similar extensions for changing attribute values do not result in clear methods for calculating the probabilities with which the different alternatives are chosen. This is a result of the fact that there is no equivalent component to the number of preference updating steps in the calculation for the probability of alternatives under internal DFT (see Equation 3.14). Instead, simulated approaches have to be considered for models utilising an internal threshold. Whilst an advantage of simulated approaches is that additional features as described in Section 2.1 can easily be added, the main advantage of external

DFT is that we can analytically calculate the expected preference values after any number of deliberation timesteps. These calculated values can then be used as the starting values for the next DFT calculation at the point where the attribute values change.

Under DFT with external thresholds, the preference values update as follows (Roe et al., 2001):

$$P_t = S \cdot P_{t-1} + V_t, \tag{6.1}$$

where P_{t-1} and P_t are column vectors representing the previous and updated preference values, V_t represents a 'valence' vector (in effect, the utility gained during one deliberation step) and S is a feedback matrix. The feedback matrix is typically specified to include decay and sensitivity parameters that allow for the impact of contextual effects such as attraction, compromise and similarity effects (Berkowitsch et al., 2014). However, for simpler cases where there are only two alternatives, we do not need to control for these effects as previous applications of DFT have found that the sensitivity parameter can becomes meaningless for choice scenarios with only two alternatives (see Chapter 4). For such cases the feedback matrix can simply act as a decay parameter:

$$S = \phi_2, \tag{6.2}$$

with $0 \leq \phi_2 \leq 1$. This additionally allows for a simpler calculation for the probabilities with which the alternatives are chosen (as defined by Equations 6.4-6.10).

Finally, the valence vector, V_t, is determined based on the attribute attended to by the decision-maker at deliberation timestep t. It is defined:

$$V_t = C \cdot M \cdot W_t + \epsilon_t, \tag{6.3}$$

where C is a contrast matrix used to rescale the attribute values M around 0. It is defined to have diagonal elements of 1 and off-diagonal elements of $-1/(x-1)$, where x is the number of attributes. W_t is a column matrix of zeros with a single 1 in the row representing the attribute that is attended to at time t. A DFT model will hence estimate a set of weights, w_a, w_b,.., w_k, with w_k giving the probability that attribute k is attended to in a given preference updating step. To allow for adjustments to the attributes M, as well as avoiding the requirement of a priori knowledge of the attributes (see Chapter 4 for details of this), we fix each of these weights to $1/x$, and estimate scaling parameters for the attributes. This allows for the application of, for example, income effects to the attributes for DFT in the same way that these effects are added to typical random utility models. Additionally, there is a noise parameter ϵ_t, which adds normally distributed values to the valence

for each alternative (with a mean of zero, independent draws and a standard deviation that is estimated).

To calculate the probabilities for the different alternatives, a DFT model simply requires the expected value of the preference vector after some number of deliberation timesteps, Φ_t, and the covariance of the preference vector Ω_t (Roe et al., 2001). With the feedback parameter S now simply represented by a constant, ϕ_2, the expectation of the preference vector after t preference updating steps becomes:

$$E[P_t] = \xi_t = \frac{1 - \phi_2^t}{1 - \phi_2} \cdot \mu + \phi_2^t \cdot P_0, \tag{6.4}$$

where P_0 is the initial preference matrix, which can be used to calculate underlying baseline preferences towards an alternative (see Chapter 4) and μ is the expectation of the valence vector, V_t.

The covariance matrix for a single set of attributes, M, also simplifies to become:

$$Cov[P_t] = \Omega_t = \frac{1 - \phi_2^{(2 \cdot t)}}{1 - \phi_2^2} \cdot \Phi, \tag{6.5}$$

where Φ is the covariance matrix for V_t (Roe et al., 2001).

However, it is worth noting that if $P_0 = 0$, t and ϕ_2 cannot be separately identified as the full set of probabilities generated by all possible combinations of t and ϕ_2 can be generated by t alone with $\phi_2 = 0$.

First, note that when $\phi_2 = 0$, the calculated values for the expectation and covariance are $\xi_t = t \cdot \mu$ and $\Omega_t = t \cdot \Phi$. Then as P_t converges to a multivariate normal distribution (Roe et al., 2001), the same set of probabilities are generated with μ and Φ or $x \cdot \mu$ and $x^2 \cdot \Phi$, where x is some constant. Consequently, the set of probabilities that are calculated with ξ_t and Ω_t are also generated by Φ and:

$$\frac{1 - \phi_2^t}{1 - \phi_2} \cdot \mu \cdot \sqrt{\left(\frac{1 - \phi_2^2}{1 - \phi_2^{(2 \cdot t)}} \right)}, \tag{6.6}$$

which can be rearranged to:

$$\sqrt{\left(\frac{(1 + \phi_2) \cdot (1 - \phi_2^t)}{(1 + \phi_2^t) \cdot (1 - \phi_2)} \right)} \cdot \mu. \tag{6.7}$$

Then, as $t > 1$ and $0 < \phi_2 \leq 1$, this is equivalent to $y \cdot \mu$, where $y \geq 1$. Consequently, we only need to estimate one parameter, y, which is equivalent to estimating $\sqrt{(t)}$ when $\phi_2 = 0$. Note that if $P_0 \neq 0$, then the value of ϕ_2 will have an impact. It is however simpler to estimate y as before, as well as some new factor, α_d, which is the factor which previous preference values

are multiplied by. Consequently, if the attribute values change (for example from M_1 to M_2) then we can define the expected preference values at the end of the deliberation process to be:

$$\xi_{2t} = t \cdot \mu_2 + \alpha_d \cdot t \cdot \mu_1 + \alpha_d^2 \cdot P_0, \tag{6.8}$$

where μ_1 and μ_2 are the expectations of the valence vectors for the different attribute matrices M_1 and M_2, and the number of preference updating steps for M_1 and M_2 are assumed to be equal.

To generalise this for R different sets of attribute matrices M_1, M_2,..., M_r, the total expected value can be calculated recursively. This results in a total expectation of:

$$\xi_{rt} = t \cdot \sum_{i=1}^{R} \left(\mu_i \cdot \alpha_d^{R-i} \right) + \alpha_d^r \cdot P_0 \tag{6.9}$$

To allow for an appropriate amount of decay for the covariance matrix relative to the expected preference, we instead multiply the covariance by α_d^2. We can then calculate the covariance matrix after R different sets of attribute matrices, for which Equation 6.5 can be expanded to:

$$\Omega_{rt} = t \cdot \sum_{i=1}^{R} \left(\Phi_i \cdot \alpha_d^{R-i} \right) \tag{6.10}$$

Once the expectation and covariance matrices are calculated for the preference vector, the probability with which the different alternatives are chosen can be calculated using multivariate normal distributions (Roe et al., 2001).

2.3 Quantum rotation models

Another alternative structure for a dynamic choice model is to use quantum rotation models. Under these models, each alternative is represented by a vector in some multidimensional Hilbert space. A 'state' vector is then used to define the probabilities of the decision-maker choosing each of the alternatives. In Chapter 5, we demonstrate that for each alternative, a 'projection length' must be estimated to generate this state vector. These projection lengths can then be used to estimate the probability with which the alternative is chosen, given that the sum of squared projection lengths must sum to one (see Section 3.1 of Chapter 5). For a binary scenario, one possibility is to consider quantum projection lengths based on trigonometric functions (Lipovetsky, 2018), where we define the probability of the first choice as:

$$Prob(Alt_1) = cos^2 \left(min \left(\frac{\pi}{2}, max\left(0, z\right) \right) \right), \tag{6.11}$$

where z corresponds to an aggregate of predictors[2]. The probability of the
second alternative is simply calculated with a sine rather than cosine function.
The functional form of z can then be based on any relevant attributes of the
alternatives to generate appropriate probabilities for each alternative being
chosen. To add a dynamic element to the model, we then require the use
of a quantum rotation. A general notation for a quantum rotation is given
by Equation 5.15 for changes in choice context. The extension to dynamic
changes in choice context can be made by specifying functional forms for
elements of the rotation matrix, M^*, based on dynamic components:

$$M^* = \begin{bmatrix} 1 & df * r_{1,2} \\ df * r_{2,1} & 1 + df \end{bmatrix}, \tag{6.12}$$

where $r_{1,2}$ and $r_{2,1}$ are estimated rotation parameters and df is based on
some dynamic function of contextual effects. Figure 6.3 gives an illustration
of how the state vector may change under dynamic choice settings. Initially,
it is of unit length. Attributes of the alternative impact the angle of the state
vector with the vectors for each of the alternatives (represented by the axes
in the figure).
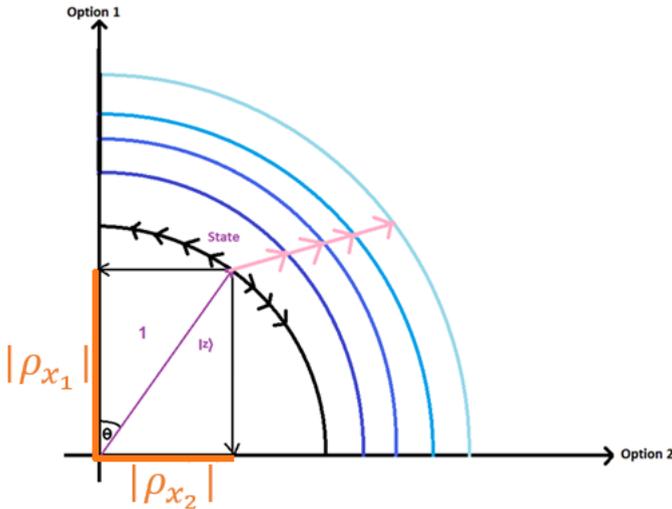
Fig. 6.3: An illustration of how the probabilities of alternatives may change under dynamic settings in quantum choice models.

The quantum rotation is then denoted by the pink arrow, which demonstrates how the choice context may impact the state vector. Crucially, as sine

---

[2]Note that the use of minimum and maximum functions here restricts $cos^2(x)$ into being a strictly monotonic function.

and cosines can be used to calculate the probability of alternatives based on the angle between the state vector and the vectors for the alternatives, the state vector does not have to remain of unit length to generate probabilities. Thus, in a dynamic choice scenario, changes in the attributes of alternatives are represented by the black arrows on the arc corresponding to circle of radius one and changes in the choice context are represented by the pink arrows which move away from this arc.

3 Empirical framework

In this section, we detail the driving dataset that is used in the empirical applications in this paper. We discuss how it is implemented for our logit, probit, quantum and DFT models, as well as describing how we add a dynamic element to logit, probit and quantum such that a fair comparison can be made with DFT (which is dynamic by nature).

3.1 Data for case study

For testing our dynamic models, we use a dataset collected on site on the southbound section of the 101 Highway in Los Angeles in June 2005 (Cambridge Systematics Inc., 2005b). It comprises of a 45 minute video recording of vehicles travelling across a 640 metre cross-section of the road, which has 5 lanes as well as an auxiliary lane which connects an on-ramp and off-ramp (see Figure 6.4).
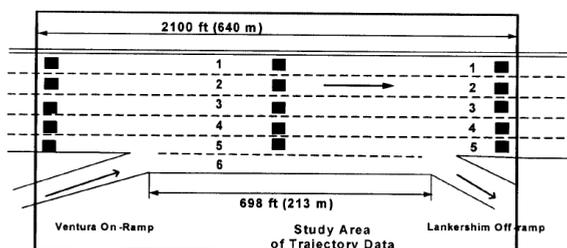
Fig. 6.4: The US-101 data collection site (Reprinted from Choudhury 2007).

This provides ideal dynamic scenarios for which naturally dynamic models such as decision field theory can be applied to. For testing the models, we look at the lane merging behaviour by the drivers who join the US-101 from the Ventura on-ramp. These drivers can join lane 5 immediately at the end of the on-ramp, or wait to merge at some point as they travel down the auxiliary lane (lane 6 in Figure 6.4). The factors impacting this decision (such as the size of the gaps on lane 5, speed and acceleration of the vehicles

189

involved, etc.) are constantly changing, as a result of both the vehicles trying
to join lane 5 and those already on it changing speed. Over the course of
the 45-minute video, 399 vehicles start on the on-ramp and merge onto lane
5. We use trajectory data which details the exact location and speeds of all
the vehicles that feature in the video (Cambridge Systematics Inc., 2005a)
at intervals of 0.1 seconds, as well as giving information on the length, width
and type of vehicle. This results in hundreds of observations for each vehicle,
with an average time of 8.73 seconds taken by a vehicle to join the US-101
from the Ventura on-ramp from the first point at which they could make the
merge.

Some of the vehicles merge immediately from the on-ramp to lane 5, whilst
others do not merge until they reach the Lankershim off-ramp. Whilst there
is likely to be significant variation in driver perception-reaction times (Fu
et al., 2016; Paschalidis et al., 2019; Wood and Zhang, 2017), we simplify
our models by assuming that drivers take a second to react to visual stimuli.
Thus, we define that the driver chooses to merge lane due to visual stimuli
at a point a second before their vehicle physically changes lane. The drivers
also have the external constraint that they must merge by at a certain point
which acts as an external threshold in the formulation. In order to estimate
the model to predict the probability that a driver will choose to merge at
a certain time step, we also require observations in which they choose not
to merge lanes. For the majority of the applications in this paper (with the
exception of results in Section 4.5), we use observations at 1 second intervals
before the moment where the driver chooses to merge (but do not consider
any moments after the driver has made the decision to merge lanes). This
results in a total of 3,293 observations across 395 vehicles.

3.2 Explanatory variables

For all the variations of models that we use, we define two utilities/preference
values, one for merging and one for staying in the current lane. Then, for
each observation in the dataset, we use four key attributes to define these
utilities/preference values:

1. Time headway in the target lane in front of the merging vehicle (in
   seconds).

2. Time headway in the target lane behind the merging vehicle (in sec-
   onds).

3. The velocity of the merging vehicle (feet per second).

4. Distance (in feet) to the point at which merging lane is no longer pos-
   sible (in this case, shortly after the driver reaches the Lankershim off
   ramp).

3.3 'Static' models

3.3.1 Logit models

In our logit models, the utility for a driver n choosing to merge at some time point t is assumed to be a function of the time headways and velocity of the merging vehicle:

$$V_{MERGE_{nt}} = U_{GF_{nt}} + U_{GB_{nt}} + U_{VEL_{nt}} + \epsilon_{nt}, \qquad (6.13)$$

where $U_{GF_{nt}}$ is the utility gained from the gap in front, $U_{GB_{nt}}$ is the utility gained from the gap behind, $U_{VEL_{nt}}$ is a utility that is dependent on the velocity of the merging vehicle and ϵ_{nt} is a random error term. Again, other specifications would be possible, such as interacting the gaps with the velocity, but this is an exploratory first step.

To calculate this utility, we then need to define specifications for the three deterministic utility components. Firstly, given that the impact of time headway is very non-linear (with increases in headway being much more important at small base values), we use a logistic transform to convert the time headway to a subjective utility[3] for the gap in front ($U_{GF_{nt}}$):

$$U_{GF_{nt}} = \beta_{GF} * \left(2 * \frac{exp(\alpha_{GF} * TH_{F_{nt}})}{1 + exp(\alpha_{GF} * TH_{F_{nt}})} - 1 \right), \qquad (6.14)$$

where β_{GF} is an estimated weight for the relative importance of the gap in front, α_{GF} is an estimated scaling parameter and $TH_{F_{nt}}$ is the time headway in front of the merging vehicle. We multiply this logistic transformation by 2 and take away 1 such that this utility can be both positive and negative. Note that some observed time headways are negative, which means that the vehicle in front in the target lane is partially adjacent to the merging vehicle (see an example of this in Figure 6.5).

As the time headways only account for the distance between the cars in one-dimension, along the length of the direction of travel (for example, in Figure 6.5, the orange arrow represents the (negative) gap between the back of the car in front, and the front of the merging car. This gap does not account for distances between the vehicles perpendicular to the direction of travel (the dashed line in Figure 6.5), thus we do not automatically translate negative time headways to extremely negative subjective values, as a vehicle may begin to merge whilst being adjacent to the vehicle in front if the relative speeds of the vehicles are very different.

The time headway in front ($TH_{F_{nt}}$) of the vehicle is defined as the gap size between the front of the merging vehicle and the back of the vehicle in

---

[3]Note that different linear and non-linear functional forms for mapping the four attributes to utilities were tested based on literature and the final functional form has been selected based on empirical testing.
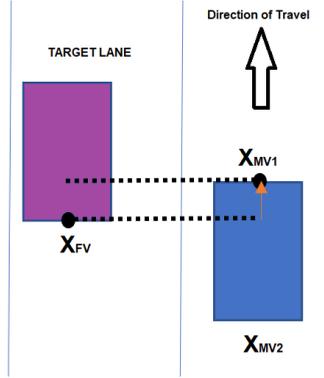
Fig. 6.5: A merging car adjacent to the vehicle in front on the target lane.

front in the target lane, with respect to the speed of the vehicle in front. Thus for decision-maker n at time point t, we have:

$$TH_{F_{nt}} = \frac{x_{VF_{nt}} - x_{MV1_{nt}}}{max(1, vel_{VF_{nt}})},$$ (6.15)

where $x_{VF_{nt}}$ is the location of the back of the vehicle in front, $x_{MV1_{nt}}$ is the location of the front of the merging vehicle and $vel_{VF_{nt}}$ is the velocity of the vehicle in front. The gap size is divided by a minimum value of 1 such that the gap is not transformed to a large value should the vehicle in front be stationary or moving extremely slowly (with a velocity of less than 1 foot per second).

We then estimate parameters β_{GB} and α_{GB} for an equivalent adjustment for the gap behind to that of the adjustment for the gap in front as defined by Equation 6.14, to get a utility U_{GB} for the gap behind. For this, the time headway behind (TH_B) the vehicle is defined as the gap size between the back of the merging vehicle and the front of the vehicle behind in the target lane, with respect to the speed of the vehicle behind:

$$TH_{B_{nt}} = \frac{x_{MV2_{nt}} - x_{VB_{nt}}}{max(1, vel_{VB_{nt}})},$$ (6.16)

where $x_{VB_{nt}}$ is the location of the front of the vehicle behind, $x_{MV2_{nt}}$ is the location of the back of the merging vehicle and $vel_{VB_{nt}}$ is the velocity of the vehicle behind.

Finally, the utility for choosing to merge also accounts for the speed of the merging vehicle:

$$U_{VEL_{nt}} = \beta_{VEL} * vel_{MV_{nt}},$$ (6.17)

with β_{VEL} an estimated parameter and $vel_{MV_{nt}}$ the velocity of the merging vehicle.

The driver may alternatively decide to stay in the same lane. We define the utility of staying in the same lane as a function of the distance traversed thus far:

$$U_{STAY_{nt}} = \alpha_{STAY} + \beta_{dist} * MV_{DE_{nt}} + \alpha_{AL} * MV_{AL_{nt}} + \alpha_{OR} * MV_{OR_{nt}},$$
(6.18)

where $MV_{DE_{nt}}$ is the distance between the merging vehicle and the last possible point at which they can merge onto the US101 and β_{dist} an estimated parameter for the relative impact of this distance. We then additionally estimate a constant α_{STAY} for choosing not to merge lanes and two further parameters α_{AL} and α_{OR}, which control for the impact of the driver being in the auxiliary lane (for which $MV_{AL_{nt}} = 1$) or on the off ramp ($MV_{OR_{nt}} = 1$).

The final specification for our logit model uses 9 estimated parameters. An assumption of type I extreme value distributions for ϵ_{nt} results in typical MNL choice probabilities.

3.3.2 Probit models

A first alternative model considered is a restricted form of probit model. With the assumption of no correlation between the error terms, a probit model also resembles a restricted form of decision field theory. A DFT model comprises of two key sources of error: the variation generated by which attribute the decision-maker attends to at each preference updating step, as well as the normally distributed error terms added to the valence vectors (see Equation 6.3). The first of these, however, becomes insignificant if the estimate for the number of deliberation timesteps is very high (which results in attribute attendance variation being averaged out). This leaves the source of error in the model coming solely from the normal error terms, the sum of which is also a normal (and hence the model becoming a restricted probit).

3.4 Dynamic specification of 'static' models

In this section we consider features for going beyond the basic models defined in the previous section. Whilst models that attempt to capture the accumulation of preference over time may provide a more 'natural' account of dynamic choice contexts, models for which the preference (or more typically, utility) for an alternative is calculated for an instantaneous moment can incorporate previous preferences through lagged variables. For example, attributes of previously chosen alternatives can be entered as a function of how similar they are to the alternatives for the current choice (Erdem, 1996), though special care is of course needed in relation to endogeneity risks. Lagged variables have also been used in the context of consumer preferences (Seetharaman, 2004) and environmental economics (Swait et al., 2004). More specifically,

the latter of these examples defines 'meta-utilities' for an alternative as a function of the product of utilities in current and previous periods. Thus dynamic discrete choice models can include parameter(s) similar to the decay parameter used in DFT, which can be adopted into our logit, probit and quantum models. The most basic approach for the incorporation of a decay parameter is to set the total utility/preference (Q_t) at a time point t as the sum of the utility/preference at time t, (P_t), combined with the utility/preference at previous time points which are subject to a decay parameter $0 \leq \alpha_d \leq 1$:

$$Q_t = \sum_{r=0}^{t-1} \left(\alpha_d^r P_{(t-r)} \right), \tag{6.19}$$

where the instantaneous-only model is obtained with $\alpha_d = 0$.

Another alternative is to use hyperbolic discounting functions (Mazur, 1987). Here we consider the generalised hyperbolic discounting function (Green et al., 1994), which results in the total utility at time t being defined:

$$Q_t = \sum_{r=0}^{t-1} \left(\frac{P_{(t-r)}}{(1 + \alpha_d \cdot r)^s} \right), \tag{6.20}$$

where s is some estimated factor for the discount rate. Examples of these discounting curves are given in Figure 6.6, with $\alpha_d = 0.5$ and $s = 1$ for the hyperbolic discounting function and $s = 2$ for the generalised hyperbolic discounting function.
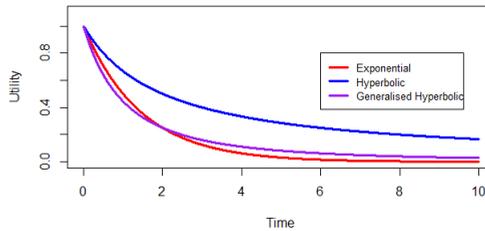
Fig. 6.6: An example of each of the discount curves tested in this paper.

Both discounting functions have been tested in dynamic choice models previously, with an exponential discounting factor used by Swait et al. (2004) and Aguirregabiria and Mira (2010) and a hyperbolic discounting function used by Fang and Wang (2015). Whilst hyperbolic functions have empirical support in a number of different fields (Green and Myerson, 1996; Kirby

and Maraković, 1995; Madden et al., 1999), examples from intertemporal choice illustrate that they do not always outperform exponential discounting functions (Read, 2001; Rubinstein, 2003).

The inclusion of these dynamic elements in the otherwise static models bridge the gap with truly dynamic models, yet do not move us away from the fact that these approaches still rely upon the stitching together of individual choices.

3.5 DFT specification

As far as the authors are aware, decision field theory in its current form has never been applied to driving behaviour, although Zhao et al. (2011) used a version 'rule-based decision field theory' in their model considering driver speed control. Whilst the choice to merge lanes is determined by the driver (and thus is more akin to a decision-maker reaching an internal threshold), there is also an external threshold in that the driver will pass a point after which merging lanes is no longer possible. Additionally, as we have trajectory data for intervals of the same length, a DFT approach as described by Equations 6.1-6.10 fits well.

A key distinction arises between DFT in estimation and in application. In application, with a given set of parameters, we can simulate the accumulation of preference until a threshold is exceeded and a decision is made, or until the decision maker runs out of deliberation time. In estimation however, we are concerned with finding values for the model parameters that explain the behaviour observed in the data. Just as with 'standard' models, this requires the maximisation of an objective function such as log-likelihood, and we need to contrast observed events with estimated probabilities. We thus need to not only estimate the probability of merging at the point where a decision is made, but similarly the probability of not merging at the time points before this event. Under DFT with an external threshold, the chosen alternative is the one with the higher preference value at some time point t. Thus, a threshold for choosing to merge can be equated to the point at which the preference for merging overtakes the preference for staying in the same lane. Hence, to make a fair comparison of DFT with alternative models, we estimate the probability that the preference value for merging is higher than the preference value for staying in the same lane after each 1 second interval. The resulting required transition from a logit to a DFT is then relatively simple.

We use attribute scaling parameters rather than attribute weights, which means that time headway in front of the vehicle, time headway behind the vehicle, the velocity of the merging vehicle and the distance (Equations 6.15 - 6.18) can be entered as four different attributes that update over time. Thus, the attribute matrix, M to be used in the DFT model can be based on the

utility functions defined previously:

$$M = \begin{bmatrix} U_{GF} & U_{GB} & U_{VEL} & 0 \\ 0 & 0 & 0 & U_{STAY} \end{bmatrix}. \tag{6.21}$$

A dynamic version of DFT with updating attribute values can then be estimated using Equations 6.1 - 6.10. As well as the same nine attribute parameters used for MNL, this model then requires three process parameters. These are the standard deviation of the valence error, the number of deliberation timesteps and a decay parameter α_d to account for the dampening of preferences when attributes update (see Equations 6.7-6.10). This allows for the current attributes for the choice as to whether to merge or not to have more impact than previous attributes, with the preferences from old time points decaying systematically over time. Note that the decay parameter ϕ_2 cannot be separately identified for our DFT model (see Section 2.2). Additionally, as we use scaling parameters, there is an overspecification due to having an additional parameter (see Section 2.3 of Chapter 4). In this case, we choose to fix the standard deviation of the error to avoid confounding. This leaves just two additional parameters.

The above use of DFT relies on the closed form calculation of probabilities, which is computationally desirable, but where the use of expectations limits the additional dynamic complexities that can be introduced. We can alternatively adopt a simulated approach, for which only Equations 6.1-6.3 apply. We have two updating preference values, one for the preference to merge, and one for the preference to stay in the same lane. The decision-maker repeatedly evaluates the attributes which results in the preference values updating over time. For some preference updating steps, the decision-maker 'samples the world' and thus the attribute matrix M in Equation 6.1 is updated. At other preference updating steps, the decision-maker does not update M, thus instead effectively 'internally resamples' the information that they already have. In this case, in the line with the other models tested, we allow updates for M to happen once per second (though of course insights from cognitive psychology could be used to inform and test different rates in future applications). We then trial different numbers of internal preference updating steps. At each preference updating step, one of the four attributes is selected randomly (with equal probability), with the preference values for merging and staying in the same lane updated in line with Equations 6.1 - 6.3. After some set number of preference updating steps, the probability of merging at each second interval can be estimated based on the mean and standard deviations for the difference between the preference values (which will also converge due to the central limit theorem), meaning that the probabilities can be calculated directly. For example, 200 simulations will give 200 estimates for the preference difference at each second interval. These values can then be used to generate a distribution which can in turn be used to

calculate the probability of choosing to merge.

3.6 Quantum model specifications

For our quantum model specifications, we do not consider the two specifications we have previously developed in Chapter 5, as these both rely on the comparison of attributes across alternatives. Given that the decision to merge or not is a binary choice, we instead use Equation 6.11. We can then define two different quantum models.

The first is 'static'. For this quantum model (QM), we set $z = \frac{\pi}{4} + U_{STAY_{nt}} - U_{MERGE_{nt}}$, such that Equations 6.13 and 6.18 can be implemented into the quantum model (where the addition of $\frac{\pi}{4}$ results in base probabilities of 0.5). Adjustments to $U_{STAY_{nt}}$ and $U_{MERGE_{nt}}$ through the use of discounting curves can then add dynamic elements, exactly as implemented for the logit and probit models.

The second variation is based on the quantum rotation model (QRM) framework described in Section 2.3. In this case, z is adjusted to $z = \frac{\pi}{4} - U_{MERGE_{nt}}$. The factors that enter $U_{MERGE_{nt}}$, which are all based on how far the decision-maker is from the end of the slip road, are instead used to represent the dynamic 'changing choice context.' Thus a rotation matrix as defined by Equation 6.12 is used, where $df = U_{MERGE_{nt}}$.

4 Results

In this section, we apply decision field theory models accounting for changing attribute values to the US-101 dataset, as well as providing comparisons based on logit, probit, quantum and quantum rotation models. We start by considering models that do not account for the dynamic nature of the dataset, which treat each choice at one second intervals as completely independent of the choices made by the same driver at later time points. We then apply both analytical and simulated versions of DFT models, comparing their performance as well as testing the impact of changing the number of attribute updates per second. Next, we apply alternative model structures based on 'static' models with variables that account for 'remembered' utilities (see Section 3.4), which adds a dynamic component to these models. Finally, we look at the inclusion of initial starting preferences, testing whether these additional parameters impact all models or just DFT models.

4.1 Basic static models

For our first set of models, we analyse the choices independently, without including any memory parameters (as detailed in Section 3.4) in any of the models. We thus test a basic version of MNL as described by Equations 6.13

and 6.18 as well as probit and quantum models, as described in Section 3.3.2 and Section 3.6. We can also test a 'static' DFT, for which the parameter α_d is fixed to a value of 0, meaning that the choices at different timepoints are not linked. For all models, we use R packages maxLik (Henningsen and Toomet, 2011) and apollo (Hess and Palma, 2019) for estimation of the log-likelihood functions.

Table 6.1: Basic static model results

Model variation	Free Parameters	Log-likelihood	BIC
Logit	9	-862.59	1798
Probit	9	-858.59	1790
Quantum	9	-856.21	1785
DFT-1	10	-858.61	1798
DFT-2	9	-858.81	1791

The performance is relatively similar across models, where the best performing model at this point is the quantum model. The estimate for the number of preference updating steps for the first DFT model (DFT-1 in Table 6.1) is very high and consequently there is no significant difference for a DFT model (DFT-2) for which the number of steps is fixed to 1000. As it has a fit that is equivalent to the probit model, it appears that there is no gain that can be attributed to stochastic variability that is generated through attention to different attributes.

For all models, we see that the size of the gap behind tends to be more important than the gap in front, with Figure 6.7 giving the contribution of U_{GF} and U_{GB} as a function of the time headways in front and behind the merging vehicle. Whilst all models appear to show very similar results, the damping in the quantum model is less extreme, with some differences between 1 and 2 seconds. This is likely a key reason for the better fit for this model.

4.2 Comparison of simulated and analytical DFT models

For our first comparison of simulated and analytical DFT models, we test the impact of fixing the number of preference updating steps. For all of the simulated models, we run 200 simulations with the use of the R package RCPP (Eddelbuettel et al., 2011). For each model, we assume that the decision-maker samples the world at one second intervals (thus attribute matrix M is updated once per second) and test rates of 1, 5, 10, 20, 50 and 100 (internal) preference updating steps per second (where for analytical DFT, this rate determines the value of t in Equations 6.9 and 6.10, and for simulated DFT, there are simply t iterative updates of Equation 6.1 per second), obtaining
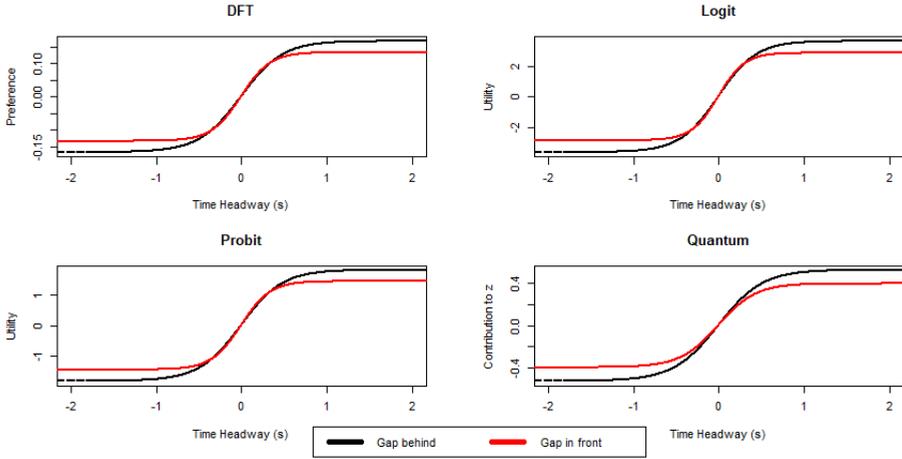
Fig. 6.7: The conversion of time headway into preferences/utilities

the results given in Table 6.2.

Table 6.2: A comparison of log-likelihoods from analytical and simulated DFT models with different numbers of preference updating steps

Preference updates per second	Sim DFT	Analytical DFT	Difference
1	-1,038.51	-1,122.82	-84.30
5	-931.00	-934.99	-3.99
10	-877.33	-883.55	-6.22
20	-860.50	-856.08	4.42
50	-851.56	-843.75	7.81
100	-843.83	-840.40	3.43

The results demonstrate that analytical DFT has a worse performance for a low number of timesteps but does better when the number of steps is high. Notably, analytical DFT is reliant on the central limit theorem (for the preference values to converge to a multivariate normal distribution) and it appears that this becomes a poor approximation when the number of steps is low. Whilst the best performance in this case is for an analytical model, these results imply that models for which the estimated number of preference updating steps is low (which was not the case for DFT-1 in Table 6.1) may perform better with a simulated version of DFT. For example, if we instead assumed that individuals sampled the world at more frequent intervals than once per second (and thus there were more frequent updates for attribute matrix M and consequently a smaller number of steps t for each

set of attribute matrices if the total number of preference updating steps per
second is to remain the same) then simulated models may be preferable. The
simulated approach also of course makes it easier to incorporate additional
dynamic elements, as discussed earlier in the paper.

An illustration of some of these models is given in Figure 6.8, which shows
200 sets of simulated preference values for one of the drivers under a simulated
DFT model (with $t = 100$).
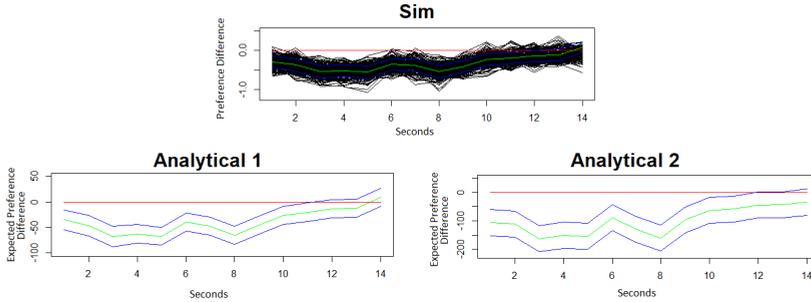
Fig. 6.8: Evolution of preference values under DFT models

The figure also shows the expected preference values for two versions
of analytical DFT, a model with a memory parameter (Analytical 1) and
one without (Analytical 2). For all models, preference differences above the
threshold of zero (which is indicated by the red horizontal line) imply that
the driver will choose to merge. The green lines give the average and ex-
pected preference values for the simulated and analytical versions of DFT
respectively. Finally, the blue lines give the standard deviations of these val-
ues (which are calculated directly for simulated DFT, and estimated based
on the covariance of the preference values for analytical DFT). For the sim-
ulated model, all 200 simulations are given. For example, we can see that
after 4 seconds not a single simulation finds a positive difference (which cor-
responds to choosing to merge). This additionally illustrates why we do not
simply evaluate the probabilities of merging through summing the number of
simulations for which the difference is positive, as this would result in zero
probabilities. Note that these probabilities are not cumulative, but are the
probability of a driver merging at that moment. Notably, the key difference
between models without a memory parameter (Analytical 2) and the other
models is that the probability of choosing to merge at the final time point
(at which point the driver chooses to merge) is lower. This is also demon-
strated in Table 6.3, which gives the average probability of observing the
chosen alternatives under three different DFT models (corresponding to the
three models in Figure 6.8), as well as the full set of probabilities for the
same driver. As a contrast, all models perform similarly for the timepoints

where the driver does not merge.

Table 6.3: Probabilities for 14 'choices' by one driver under four different DFT models

Model	Simulated	Analytical 1	Analytical 2
deliberations per second	100	100	1000
α_d	0.35	0.25	0.00
Log-likelihood	-843.83	-840.40	-858.81
Average stay probability	0.9075	0.9077	0.9061
Average merge probability	0.3245	0.3230	0.3057
Seconds 1	0.9693	0.9627	0.9890
2	0.9886	0.9887	0.9920
3	0.9997	0.9996	0.9998
4	0.9996	0.9997	0.9995
5	0.9997	0.9999	0.9997
6	0.9921	0.9877	0.9741
7	0.9892	0.9961	0.9978
8	0.9995	0.9999	0.9998
9	0.9979	0.9951	0.9827
10	0.9323	0.9323	0.9173
11	0.9140	0.8672	0.8974
12	0.8310	0.7747	0.8312
13	0.7847	0.7449	0.8279
14	0.2611	0.3107	0.2272

4.3 Models accounting for decay parameters

Whilst DFT provides a natural method for the evolution of preferences over time, models that calculate probabilities based on a single utility calculation can also capture the impact of previous preferences through decay parameters (see Equations 6.19 and 6.20). This results in a significant improvement in model fit for all of the models, with the log-likelihoods of models implementing an exponential discounting function given in Table 6.4, and analytical DFT models (with free timestep parameters) also displayed for comparison. We simplify all models by assuming that the unobserved errors are independent over time.

For the models with and without a decay parameter, there is very little difference between the different types of models, although the quantum model is no longer the best performing model when a decay is included. Additionally, there is now also a small difference between DFT and probit, although this difference is not significant. The estimate for the decay parameter is slightly higher for DFT than the other models. Table 6.4 also gives the ratio β_{GF}/β_{GB}, which gives the estimated relative importance of the gap

Table 6.4: Performance and model outputs for models with basic decay parameters

| | Without memory | With memory | | |
	Log-likelihood	Log-likelihood	α_d estimate	β_{GF}/β_{GB}
Logit	-862.59	-840.36	0.227	0.732
Probit	-858.59	-838.29	0.237	0.755
Quantum	-856.21	-837.67	0.235	0.775
DFT	-858.81	-837.48	0.250	0.759

in front of the vehicle compared to the gap behind the vehicle. The results here demonstrate that all models find that the gap behind is more important, with the quantum model assigning the highest importance to the gap in front relative to the other models.

Next, we consider improving the models by increasing the flexibility of the decay parameters. For all models, we consider hyperbolic discounting functions as well as utilising different discounting parameters for the utility to stay and the utility to merge. The results of these models are given in Table 6.5. For the hyperbolic functions, we consider two variations. The first of these is based on the original hyperbolic discounting function (for which s is fixed to 1 in Equation 6.20). The second uses both hyperbolic discounting parameters.

Three clear patterns emerge from these models. Firstly, quantum models perform slightly but not significantly better than probit models, but significantly better than logit models. Secondly, additional parameters allowing for different decay rates for the utility for merging compared to the utility for staying in the same lane significantly improve model fit. Finally, exponential decay models perform better than hyperbolic decay models, but worse than generalised hyperbolic decay models (for which s in Equation 6.20 is a free parameter). The shape of the discount curves estimated are given for the best exponential and best hyperbolic versions of each model in Figure 6.9.

Notably, the same patterns emerge for all of the different models. There is no difference in the decay for the utility to merge, regardless of whether hyperbolic or exponential discounting decays are used, but there is a substantial difference for the decay for the utility to stay. This implies that the gain in model fit found by the hyperbolic decay models is possibly due to the increased flexibility that allows for an initially sharp decay before a more gradual decrease after 1 second (the purple lines in Figure 6.9. Furthermore, Figure **??**, which gives the evolution of utilities under logit, probit and quantum models (for the same driver whose preferences are illustrated in Figure 6.8), suggests that the addition of decay parameters appears to decrease the probability of choosing to merge. As before with DFT, the models predict a greater probability of choosing to merge at the point where the driver does

Table 6.5: Log-likelihoods from models utilising additional decay parameters

Discount function	Exponential	Exponential	Hyperbolic	Hyperbolic	Hyperbolic
Memory parameters	1	2	1	2	4
Equivalent decay	yes	no	yes	no	no
Logit	-840.36	-834.91	-847.05	-840.57	-830.45
Probit	-838.29	-831.97	-843.81	-838.46	-827.78
Quantum	-837.67	-830.53	-842.13	-837.93	-828.73
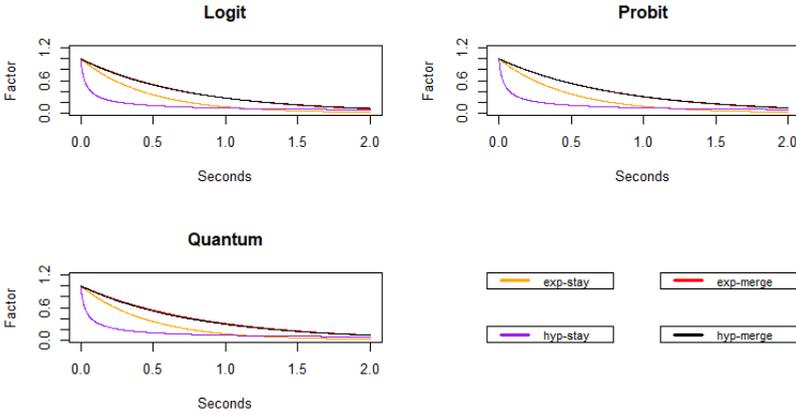
Fig. 6.9: The discount curves estimated by the different models

actually choose to merge.
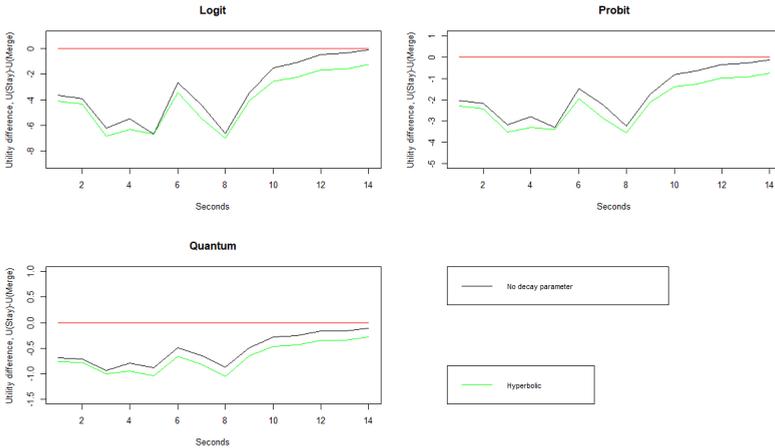
Fig. 6.10: Evolution of utilities under logit, probit and quantum models

4.4 Inclusion of initial preferences

Thus far, the alternative specific constant that we have utilised in our models has been applied for every new set of attribute values. Whilst this parameter helps capture the baseline preferences for alternatives (note that the drivers choose to merge at only 12% of the timepoints), it does not capture a baseline preference at the start of the decision process. For example, a driver may

initially be keen to merge lanes, but then quickly realise that it is not possible upon approaching the motorway. For DFT models, this initial preference can easily be captured by estimating P_0 in Equation 6.1. For the other models, this initial preference can be captured through the addition of a constant in one of the utilities/projection lengths (that is applied at the first time point only). The results of adding this parameter are given in Table 6.6. The logit, probit and quantum models without the initial preference parameters are based on the best fitting models displayed in Table 6.5. We also give the results for quantum rotation models, where there are no memory parameters, but dynamic effects are captured through the incorporation of quantum rotation matrices. Finally, for analytical DFT, the model without an initial preference parameter has the number of deliberation timesteps fixed to 1000.

Table 6.6: The impact of adding an initial preference parameter on the log-likelihood for the different models

Model	Decay Type	without	+ P0	Difference
Simulated DFT	DFT	-840.76	-824.57	16.19
Analytical DFT	DFT	-837.48	-831.44	6.04
Probit	Exp	-831.97	-831.39	0.58
Probit	Hyp	-827.78	-827.31	0.47
Logit	Exp	-834.91	-834.34	0.57
Logit	Hyp	-830.45	-830.42	0.03
Quantum	Exp	-830.53	-829.59	0.94
Quantum	Hyp	-828.73	-828.67	0.06
Quantum Rotation	Rotation	-836.53	-835.46	1.07

Crucially, DFT is the only model that benefits from the addition of an initial preference parameter. It is also worth noting that a simulated version of DFT improves significantly more than an analytical version of DFT. This implies that some of the benefit of an initial parameter is averaged out by an analytical DFT model.

4.5   Impact of information update rate

Whilst all of the results thus far have only used observations at 1 second intervals before the moment where the driver chooses to merge, we also consider models applied to the data where more observations are included. This results in different frequencies of 'visual information updating speeds' for the decision-makers (i.e. different rates for how often the attributes of alternatives are updated). The key benefit of testing across different rates is that it allows us to test how the estimate for the number of preference updating

steps in a DFT model is impacted by the information updating rate. This
is a result of the fact that under our implementation, it is assumed that an
individual considers each set of attribute values for the same number of pref-
erence updating steps. We estimate full (analytical) DFT models on datasets
that include 1, 2, 5 and 10 data updates per second. The results of these
models are given in Table 6.7.

Table 6.7: Results from DFT models applied with different visual update rates

Data Updates per second	1	2	5	10
Observations	3,293	6,391	15,718	31,245
Log-likelihood	-831.75	-1,110.15	-1,473.87	-1,751.41
t estimate:	347	202	114	43

Crucially, the estimate for the number of preference updating steps per
visual update decreases as the of data updating steps increases, resulting in
approximately the same total of preference updating steps per second. A
visual representation of this result is given in Figure 6.11.
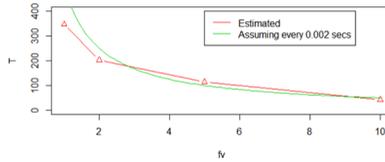
Fig. 6.11: Relationship between the frequency of visual (fv) and mental updating steps
(T)

Additionally, we also compare the results of our final models across the
dataset with two different assumptions for the visual update rate. Table 6.8
gives the log-likelihood of models where 0.1 and 1 second intervals are used
for data update rates (with exponential discount rates used for the 'static'
models).

The relative performance of the models changes depending on which up-
date rate is used. Whilst the quantum rotation model has the worse fit than
the quantum model for a 1 second interval data, it has better fit for 0.1 sec-
ond interval data. Similarly, probit performs relatively better for 0.1 second
data, obtaining the second best model fit. Across both datasets, however,
DFT is the best performing model.

Table 6.8: Performance of the different models for different data update frequencies

	pars	0.1s interval		1s interval	
		LL	BIC	LL	BIC
Logit	12	-1,768.58	3,661.36	-830.42	1,785.04
Probit	12	-1,753.21	3,630.61	-831.39	1,786.98
DFT	12	-1,750.80	3,625.79	-824.57	1,773.34
Quantum	12	-1,759.13	3,642.46	-829.59	1,783.38
Quantum Rotation	14	-1,756.04	3,656.98	-835.46	1,815.81

5 Conclusions and future steps

This work provides a first step towards bridging the gap between mathematical psychology and choice modelling for dynamic data, for which the attributes for alternatives change over time. In particular, we focus on the implementation of a dynamic decision field theory model for such data. As DFT models already attempt to capture the deliberation process, the inclusion of attribute changes whilst the decision-maker deliberates on their alternatives does not imply fundamental changes to the mathematical structure of the model, but has never before been implemented, as far as the authors are aware. In the context of driving decisions, we find that a dynamic DFT model can capture the choice of whether to merge lanes or not. In addition, a simulated DFT model can account for initial preferences, which do not appear to be picked up by alternative model structures.

The work conducted in this paper, however, is just a first step, as there are many more features that a model for dynamic data could include. Given that the data tested here is based on video recorded data, we have not included any sociodemographics or individual-specific attributes. It is easily possible, for example, that a driver's reaction times and risk propensity could have a significant impact on the choice of whether to accept a gap or not. Additionally, an individual's speed at which they process new information will likely impact how they perceive the new information, with previous results suggesting that there is a strong delay effect on the processing of changing information (Holmes et al., 2016). Furthermore, the separation of processing and action times will likely impact models. There is however, a clear additional advantage of a simulated DFT model, in that the addition of such features to control for heterogeneous driving behaviour can easily be accommodated, as adding distributions for new parameters will not result in longer model runtimes. These further steps would help create a *truly* dynamic model, as at present DFT operates similarly to the 'static' models in that it is reliant on the decay parameter α_d.

Additionally, a decision field theory model for changing attributes could

also be compared to alternative model structures based on alternative accumulator models. For example, the linear ballistic accumulator model could easily be adapted to include drift rates that are applied for some amount of time before they are updated when the attributes change.

Furthermore, whilst the emphasis of this paper is not on the driving behaviour itself, future work could compare the models developed in this paper with models traditionally used for gap acceptance and lane merging tasks, for which a large array of models have previously been specifically developed (Brilon et al., 1999). Additionally, further tests of the models developed in this paper should consider transferability to other driving behaviour datasets, for which logit models already perform well (Rossi et al., 2013). A number of other factors that influence merging behaviour could also be added, with, for example, neither vehicle accelerations nor traffic conditions taken into account in the models specified in this paper. The work in this paper also assumes that the choice to merge or not is a binary decision. An alternative approach would be to attempt to capture the distance travelled in the direction of the target lane at each time interval, as the action of merging can take a few seconds and a driver may start merging before changing their mind. Accumulator models accounting for changing attributes may also look very different for differing choice contexts. In the dataset in this paper, the pattern of choices is always the same (the driver 'chooses' not to merge for some number of time intervals before 'choosing' to merge in the final time interval). Thus the key impact of the decay parameters in these models may simply be accounting for this fact. More interesting behavioural insights may be generated by a more 'random' pattern of choices, for which the impact of decay parameters may be very different. The dynamic logit, probit and quantum models in this paper also rely on the assumption that there is no correlation in the unobserved factors over time. Whilst this is an approach that has previously been applied to a probit model with lagged variables (Papatla and Krishnamurthi, 1992), and this approach may be suitable for larger time intervals, it becomes less likely that this assumption is necessarily realistic nor valid for shorter time intervals of, for example, 0.1 seconds. Further work should consider methods for treating possible correlations of unobserved factors over time.

Overall, the work in this paper suggests that accumulator models that account for changing attributes could provide a useful tool for the study of rapidly changing choice contexts, demonstrating that there is clearly extensive scope for future developments of such models.

Acknowledgements

References

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67.

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.

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

Bjørner, T. B. and Leth-Petersen, S. (2005). Dynamic models of car ownership at the household level. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, pages 57–75.

Boxall, P., Adamowicz, W. L., and Moon, A. (2009). Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Australian Journal of Agricultural and Resource Economics*, 53(4):503–519.

Brilon, W., Koenig, R., and Troutbeck, R. J. (1999). Useful estimation procedures for critical gaps. *Transportation Research Part A: Policy and Practice*, 33(3-4):161–186.

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

Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.

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.

Cambridge Systematics Inc. (2005a). NGSIM US 101 Data Analysis. http://www.ngsim.fhwa.dot.gov.

Cambridge Systematics Inc. (2005b). NGSIM US 101 Data Analysis: Summary Report. *Prepared for Federal Highway Administration.*

Choudhury, C. F. (2007). *Modeling driving decisions with latent plans.* PhD thesis, Massachusetts Institute of Technology.

Cirillo, C. and Xu, R. (2011). Dynamic discrete choice models for transportation. *Transport Reviews*, 31(4):473–494.

Diederich, A. and Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, 47(3):304–322.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628.

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.

Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing science*, 15(4):359–378.

Fang, H. and Wang, Y. (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *International Economic Review*, 56(2):565–596.

Fosgerau, M., Frejinger, E., and Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56:70–80.

Fu, C., Zhang, Y., Bie, Y., and Hu, L. (2016). Comparative analysis of driver's brake perception-reaction time at signalized intersections with and without countdown timer using parametric duration models. *Accident Analysis & Prevention*, 95:448–460.

González-Valdés, F. and de Dios Ortúzar, J. (2018). The stochastic satisficing model: A bounded rationality discrete choice model. *Journal of choice modelling*, 27:74–87.

Green, L., Fry, A. F., and Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological science*, 5(1):33–36.

Green, L. and Myerson, J. (1996). Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting time. *American Zoologist*, 36(4):496–505.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., and Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, 35(6):2476–2484.

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

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

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.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2016). The probit choice model under sequential search with an application to online retailing. *Management Science*, 63(11):3911–3929.

Kirby, K. N. and Maraković, N. N. (1995). Modeling myopic decisions: Evidence for hyperbolic delay-discounting within subjects and amounts. *Organizational Behavior and Human decision processes*, 64(1):22–30.

Kitamura, R. and Bunch, D. S. (1990). Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components.

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.

Lipovetsky, S. (2018). Quantum paradigm of probability amplitude and complex utility in entangled discrete choice modeling. *Journal of choice modelling*, 27:62–73.

Madden, G. J., Bickel, W. K., and Jacobs, E. A. (1999). Discounting of delayed rewards in opioid-dependent outpatients: exponential or hyperbolic discounting functions? *Experimental and clinical psychopharmacology*, 7(3):284.

Mai, T., Fosgerau, M., and Frejinger, E. (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75:100–112.

Marczak, F., Daamen, W., and Buisson, C. (2013). Key variables of merging behaviour: empirical comparison between two sites and assessment of gap acceptance theory. *Procedia-Social and Behavioral Sciences*, 80:678–697.

References

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, pages 55–73.

Noguchi, T. and Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological review*, 125(4):512.

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.

Papatla, P. and Krishnamurthi, L. (1992). A probit model of choice dynamics. *Marketing Science*, 11(2):189–206.

Paschalidis, E., Choudhury, C. F., and Hess, S. (2018). Modelling the effects of stress on gap-acceptance decisions combining data from driving simulator and physiological sensors. *Transportation research part F: traffic psychology and behaviour*, 59:418–435.

Paschalidis, E., Choudhury, C. F., and Hess, S. (2019). Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour. *Analytic Methods in Accident Research*, 22:100089.

Polus, A., Shiftan, Y., and Shmueli-Lazar, S. (2005). Evaluation of the waiting-time effect on critical gaps at roundabouts by a logit model. *European Journal of Transport and Infrastructure Research*, 5(1):1–12.

Ratcliff, R. and Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human perception and performance*, 26(1):127.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of risk and uncertainty*, 23(1):5–32.

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.

Rossi, R., Meneguzzer, C., and Gastaldi, M. (2013). Transfer and updating of logit models of gap-acceptance and their operational implications. *Transportation Research Part C: Emerging Technologies*, 28:142–154.

Rubinstein, A. (2003). "economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216.

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143.

Rysman, M., Gowrisankaran, G., et al. (2009). *Dynamics of Consumer Demand for New Durable Goods*. National Bureau of Economic Research.

Schiraldi, P. (2011). Automobile replacement: a dynamic structural approach. *The RAND journal of economics*, 42(2):266–291.

Seetharaman, P. (2004). Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science*, 23(2):263–271.

Swait, J., Adamowicz, W., and van Bueren, M. (2004). Choice and temporal welfare impacts: incorporating history into discrete choice models. *Journal of Environmental Economics and Management*, 47(1):94–116.

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. and Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive science*, 35(8):1518–1552.

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multi-attribute choice. *Psychological Review*, 117(4):1275.

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

Voskuilen, C., Ratcliff, R., and Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, 73:59–79.

Wang, Z., Solloway, T., Shiffrin, R. M., and Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26):9431–9436.

Wood, J. and Zhang, S. (2017). Evaluating relationships between perception-reaction times, emergency deceleration rates, and crash outcomes using naturalistic driving data. Technical report, Mountain-Plains Consortium, North Dakota State University, Fargo, ND.

Zhang, S., Lee, M. D., Vandekerckhove, J., Maris, G., and Wagenmakers, E.-J. (2014). Time-varying boundaries for diffusion models of decision making and response time. *Frontiers in Psychology*, 5:1364.

References

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 Sage CA: Los Angeles, CA.

Chapter 7

Improving forecasts and behavioural insights by applying model averaging across multiple choice models

Thomas Hancock[1], Stephane Hess[1], Andrew Daly[1] & James Fox[2]

Abstract

Despite the frequent use of model averaging in many disciplines from weather forecasting to health outcomes, it is not yet an idea often considered in travel behaviour or choice modelling. The idea behind model averaging is that a single model can be created by calculating contribution weights for a set of candidate models, depending on their relative performance, thus creating an 'average'. In this paper, we demonstrate that this idea can be used effectively for two key alternative purposes within travel behaviour modelling. The first is to apply model averaging across a number of models with the aim of improving model fits. We demonstrate that model averaging can be effective in applications across multiple complex models as well as across simpler models for large-scale choice datasets, both in estimation and in forecasting with subsets of validation samples. The second key area of application is to contrast model averaging with latent class models (LCMs). LCMs have traditionally been used for taste heterogeneity, but are increasingly used as a tool for capturing heterogeneity in other components, such as information/attribute processing and decision rules. This often leads to substantial improvement in model fit and the apparent finding of large clusters of individuals making choices in ways that are substantially different from those used by others. Our results however demonstrate that model averaging leads to significant reductions in the amount of heterogeneity of the type analysts have sought to

[1]Choice Modelling Centre and Institute for Transport Studies, University of Leeds (UK)
[2]RAND Europe

uncover with latent class structures of late.

1 Introduction

Whilst there have thus far been very limited applications in choice modelling using model averaging, it is a popular method elsewhere with Bayesian model averaging being used regularly in medical statistics (Hoeting et al., 1999), ecology (Wintle et al., 2003) and biology (Posada and Buckley, 2004). Additionally, ensembles are often used to combine neural networks (Gazder and Ratrout, 2015; Moretti et al., 2015). Typically, model averaging or ensemble methods can be used to allow a modeller to establish a single model by calculating relative contribution weights for a set of candidate models. Within health, Bayesian model averaging has been successfully used to improve the prediction of who is at risk of a stroke (Volinsky et al., 1997), at risk of a coronary event (Wang et al., 2004) and to understand the relation between arsenic levels and cancer rates (Morales et al., 2006). Additionally, model averaging is often used for pooling forecasts from different models. This is particularly common for meteorological forecasting, with model averaging having been used to predict the surface temperature of the ocean (Raftery et al., 2005) and also wind speeds (Sloughter et al., 2010). It is also used in other fields for tasks such as predicting levels of economic inflation (Wright, 2009).

Choice modellers, by contrast, may often consider a set of candidate models, detail the advantages and disadvantages of each, but then subjectively choose only one model to use in the main application or reporting of key results. Consequently, it seems surprising that model averaging has not yet made the transition into mainstream choice modelling given that it can capture the benefits from a number of models and combine them into one model. This is the first key application of model averaging that we test in this paper: averaging across a large number of candidate models as well as looking at the impacts on important model outputs such as cost and time elasticities. This is one reason for applying model averaging in such a way, i.e. the case where multiple candidate models all have advantages and disadvantages and there is no clear cut case for choosing which is best. Another rather different context arises in the case of very large-scale applications, either with large datasets or large choice sets, where we may not always be able to use as complex a model as we might otherwise choose to use due to the computational running time of complex models. This reason for model averaging will become even more timely in the context of increasing reliance on big data.

The second key application of model averaging is to consider insights generated through the comparison of latent class models and model averaging. Latent class structures have long been used as a tool for introducing het-

erogeneity across individual decision makers in choice models (Greene and Hensher, 2003; Hess, 2014). Over the last decade, there has also been increasing interest in using the models to allow for heterogeneity in the actual underlying model structure across individuals, with two key applications, in decision rule heterogeneity and in information processing work. While the former has received more attention, the latter work actually takes historical precedence.

A key interest in the field of information processing strategies (IPS) or attribute processing strategies (APS) has been the notion that some decision makers may actually make their choices based on only a subset of the attributes that describe the alternatives at hand. This phenomenon is typically referred to as attribute non-attendance (ANA) or attribute ignoring, and an in-depth review of work in this area is given in Hensher (2010). The interest in this topic in the present discussions comes in the context of ways to accommodate ANA in models. A key role in this area was played by the early discussions in Hess and Rose (2007), who proposed the use of a latent class approach to accommodate ANA, a method since adopted by numerous other studies (e.g. Campbell et al., 2010; Hensher and Greene, 2010; Hensher et al., 2012; Hole, 2011; Scarpa et al., 2009). With this approach, different latent classes relate to different combinations of attendance and non-attendance across attributes. For each attribute treated in this manner, there exists a non-zero coefficient (to be estimated), which is used in the *attendance classes*, while the attribute is not employed in the *non-attendance* classes, i.e. the coefficient is set to zero. In a complete specification, covering all possible combinations, this would thus lead to 2^K classes, with K being the number of attributes, where a given coefficient will take the same value in all classes where that attribute is included.

In addition to the vector β, we now have a SxK matrix Λ, in which each row contains a different combination of 0 and 1 elements, where $S = 2^K$. Next, let $A \circ B$ be the element-by-element product of two equally sized vectors A and B, yielding a vector C of the same size, where the kth element of C is obtained by multiplying the kth element of A with the kth element of B. Using this notation, the specific values used for the taste coefficients in class s are then given by the vector $\beta_s = \beta \circ \Lambda_s$. The likelihood for decision maker n is then given by:

$$L_n\left(\beta, \pi\right) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} P_{ni^*t}\left(\beta_s = \beta \circ \Lambda_s\right). \tag{7.1}$$

A different application of such heterogeneous structures in different classes has arisen in the context of decision rule heterogeneity. There has long been interest in the notion that different individuals make their decisions in different ways, going back to work in psychology in the 1970s (Montgomery

and Svenson, 1976). Although structures belonging to the family of random utility models have come to dominate, it is important to recognise that alternative paradigms for decision making have been proposed, for example the elimination by aspects model of Tversky (1972), but also more recent work based on the concepts of happiness (Abou-Zeid and Ben-Akiva, 2010) and regret (Chorus, 2010; Chorus et al., 2008). The evidence in the literature is that which paradigm works best is very much dataset specific. Hess et al. (2012) put forward the hypothesis that variations in decision rules may be across decision makers with a single dataset, not just across datasets, and propose the use of a confirmatory latent class approach in this context.

Specifically, let $L_n (\beta_m, m)$ give the probability of the observed sequence of choices for decision maker n, conditional on using a choice model identified as m, where this uses a vector of parameters β_m. The Hess et al. (2012) framework is based on the idea that M different behavioural processes are used in the data. The probability for the sequence of choices observed for decision maker n is now given by:

$$L_n (\beta, \pi) = \sum_{m=1}^{M} \pi_{n,m} L_n (\beta_m, m), \qquad (7.2)$$

where we use different behavioural processes in different classes, with the probability of decision rule class m for decision maker n given by $\pi_{n,m}$. Hess et al. (2012) additionally allow for random heterogeneity in parameters within individual decision rule classes, such that:

$$L_n (\Omega, \pi) = \sum_{m=1}^{M} \pi_{n,m} \int_{\beta_m} L_n (\beta_m, m) f (\beta_m, \Omega_m) \, \mathrm{d}\beta_m, \qquad (7.3)$$

where $\beta_m \sim f (\beta_m, \Omega_m)$ and $\Omega_m = \langle \Omega_1, \ldots, \Omega_M \rangle$.

Hess et al. (2012) use the model to allow for mixtures between random utility maximisation, random regret minimisation and elimination by aspects. In later work, Hess and Stathopoulos (2012) use an approach as in Walker and Ben-Akiva (2002) and Hess et al. (2013a), making the class allocation a function of a latent factor, which in this case also explains decision makers' real world choices.

At this stage, it should be noted that a latent class model mixing various decision rules is just one example of a wider set of structures that combine different models. A further possibility for example would be a model using different GEV nesting structures in different latent classes, somewhat similar in aims to the work of Ishaq et al. (2013). Finally, a separate body of work looks at using different choice sets in different classes, in the context of choice set generation work (see e.g. Ben-Akiva and Boccara 1995; Swait and Ben-Akiva 1985 and Gopinath 1995, section 2.7).

While the work using latent class structures for heterogeneity in either decision rules or information processing strategies has been shown to lead to substantial improvement in fit and apparent meaningful insights (see references above), it has also not been without criticism. In particular, concerns have been raised about extensive risk of confounding between taste heterogeneity and heterogeneity in the process or model structure. In a traditional latent class model, the different β parameters in different classes are used solely to uncover taste heterogeneity. In a latent class model that combines different structures in different classes, these individual models will themselves be making use of different β parameters, while in the case of ANA, they will use different combinations of the β parameters. There is then the real possibility that evidence of a substantial class allocation probability for different classes will be driven by heterogeneity in sensitivities rather than actual process. These concerns have found empirical support in the work of Hess et al. (2013b) who show that the share for non-attendance classes reduces substantially when allowing for additional random heterogeneity, while the work of Hess et al. (2016) shows that allowing for random heterogeneity in the parameters of RUM and RRM models within a RUM-RRM mixture model substantially reduces the extent of decision rule heterogeneity.

The use in practice of such latent class models allowing for different structures in different classes continues to be very popular (Boeri and Longo, 2017; Dey et al., 2018) despite these concerns. A key reason is likely that the inclusion of additional taste heterogeneity, as in the work of Hess et al. (2013b) and Hess et al. (2016) is computationally very difficult. In the present paper, we thus use a different approach by highlighting how model averaging can be used as a diagnostic tool for the potential confounding between taste heterogeneity and other heterogeneity.

Model averaging, in this context, can be implemented as a sequential latent class model. Whereas a fully flexible model simultaneously estimates the parameters of the class component models as well as the class shares, a model averaging approach estimates the separate classes as individual models first, before estimating the class shares separately with the individual model parameters fixed. Thus the second key aim of using model averaging in this paper is to investigate potential cases of confounding in models using simultaneous estimation of different model structures. Of course, a caveat applies in that it is also possible that the presence of decision rule heterogeneity and/or heterogeneity in processing strategies can only be uncovered when estimating models in which the parameter estimates for the different subclasses are informed more by some individuals in the data than by others, as would be the case in simultaneous estimation.

The remainder of this paper is organised as follows. First, we present a methodology section demonstrating how we apply model averaging and how to get outputs such as elasticities from model averaging. Next, we present our

empirical applications, where we use model averaging to improve model fit
in both estimation and forecasting and to obtain welfare measures averaged
across a set of candidate models. This is followed by our work on attribute
non-attendance and decision rule heterogeneity. The final section summarises
our findings and presents directions for future research.

2 Methodology

2.1 Model averaging in estimation

To apply model averaging, we first estimate a number of different individual
models, where say $L\left(C_n \mid m, \Omega_m\right)$ gives the likelihood of the sequence of
choices C_n observed for person n, conditional on using model m, where this
model uses a vector of parameters Ω_m. We have that:

$$L\left(C_n \mid m, \Omega_m\right) = \int_{\beta_m} \prod_{t=1}^{T_n} P_m\left(j_{n,t}^* \mid \beta_m\right) f_m\left(\beta_m \mid \Omega_m\right) \mathrm{d}\beta_m. \tag{7.4}$$

In this general notation, we have that $P_m\left(j_{n,t}^* \mid \beta_m\right)$ gives the probability
of the observed choice $j_{n,t}^*$ for decision maker n in choice situation t, con-
ditional on using model m, where we allow for a general notation such that
the parameters β_m are distributed according to $f_m\left(\beta_m \mid \Omega_m\right)$. Of course, it
is possible that no random heterogeneity is used in which case the integral
drops out, or that a latent class structure is used, replacing the integral with
a weighted summation.

   An analyst will estimate M different such individual models, of differing
form, each yielding a set of parameters and a likelihood at the individual
level $L\left(C_n \mid m, \Omega_m\right)$. The set of M models could combine models using
different distributions for random parameters, models with different socio-
demographic specifications, models of different types, different specifications
for IPS or different specifications in terms of underlying decision rule. The
model averaging process then computes the overall likelihood for person n as
the weighted average across M models, with:

$$L_n\left(\pi_n, \Omega\right) = \sum_{m=1}^{M} \pi_{m,n} L\left(C_n \mid m, \Omega_m\right), \tag{7.5}$$

where $\sum_{m=1}^{M} \pi_{m,n} = 1$ and $0 \leq \pi_{m,n} \leq = 1$. This overall likelihood is
conditional on the vector of weights $\pi_n = \langle \pi_{1,n}, \ldots, \pi_{M,n} \rangle$ and the combined
parameter estimates from the different models $\Omega = \langle \Omega_1, \ldots, \Omega_M \rangle$.

Of course, this structure takes the form of a latent class model, but two core
differences apply.

Firstly, in a latent class model, an analyst simultaneously estimates the class allocation probabilities and the within class probabilities. In model averaging, individual models are estimated for the entire sample, and then the weights for these models are estimated, conditional on the parameters obtained during the individual model estimations. Model averaging is thus a sequential rather than simultaneous process. This is clearly computationally much easier, but also in fact allows a situation where the individual models come from different teams of analysts. In fact, the estimation of the weights in Equation 7.5 does not require the parameters of the individual models, or even the mathematical formulation of the probabilities for individual models, but simply relies on the person-specific likelihoods obtained with the individual models. Model averaging will almost inevitably lead to a lower model fit than the estimation of a simultaneous structure, but of course the general situation is one where this simultaneous structure is not feasible to be estimated.

Secondly, in a latent class model, it is generally the case that the same overall model structure is used in different classes, though this is by no means necessary (cf. Hess et al., 2012). In model averaging, a different model specification, in terms of model structure and/or e.g. utility specification, is required for the different models as the separate estimation of the same structure for different m would of course yield the same fit and parameter estimates.

Just as in standard latent class approaches, it is entirely possible to specify a class allocation model for the model weights, i.e. making π_n a function of characteristics of the individual n.

2.2 Model averaging in application

The use of model averaging produces a new likelihood at the person level, $L_n(\pi_n, \Omega)$, where the overall model averaging log-likelihood (across N people) is given by:

$$LL(\Omega, \pi) = \sum_{n=1}^{N} log\left(L_n(\pi_n, \Omega)\right). \tag{7.6}$$

This overall log-likelihood will be at least as good as the log-likelihood for the best model out of the set of M different models. However, model fit alone is not the key reason for model averaging[3]. The output from the model averaging estimation process is a vector of weights for different models, where these are potentially individual specific, i.e. $\pi_{m,n}$ for person n and model m. These weights can then be used in model application, with two key uses.

Firstly, let $P_n(j \mid S, m, \Omega_m)$ give the probability of individual n choosing a specific alternative j out of a choice set S, conditional on model m, where

---

[3]For a full discussion of different model averaging methodologies and benefits, readers should refer to Claeskens (2008).

this probability may again require integration. We can then compute the probability of this alternative j under model averaging as:

$$P_n \left(j \mid S, \pi_n, \Omega \right) = \sum_{m=1}^{M} \pi_{m,n} P_n \left(j \mid S, m, \Omega_m \right), \qquad (7.7)$$

and an analyst can then for example use these weighted predictions in sample enumeration or other forecasting.

Secondly, let $W_n \left(j \mid S, m, \Omega_m \right)$ be some model output for individual n and choice set S, conditional on model m. This could for example be a willingness-to-pay (WTP) measure from model m or an elasticity measure. It is then similarly possible to compute a model average version of this output, using:

$$W_n \left(j \mid S, \pi_n, \Omega \right) = \sum_{m=1}^{M} \pi_{m,n} W_n \left(j \mid S, m, \Omega_m \right), \qquad (7.8)$$

Any measures such as WTP or elasticities thus need to first be calculated for the individual models before being averaged across models. The calculation will likely differ across models and may involve simulation for some of the models if they incorporate random heterogeneity. If this is the case, it is advisable to use the entire distributions in model averaging rather than just relying on the moments from individual models if some non-normal distributions are included.

The key advantage of this process is that the calculation of these predictions or derived measures is informed by the results of a number of different models, and is thus potentially more robust to mis-specification of the individual models. It is similarly possible to compute variances for the outputs of Equation 7.6 and 7.8, though we rely just on the mean outputs in the present paper.

3 Empirical application

In this section, we first give details on the different datasets used in this paper. We then test the two key applications of model averaging. First, for each dataset, we demonstrate how model averaging can improve model fit, as well as considering elasticities and willingness-to-pay outputs from model averaging. Second, we contrast model averaging to latent class models to generate insights for information processing work and also decision rule heterogeneity work.

3.1 Data

We use three different datasets for trialling model averaging. The first is a typical SP dataset with the latter two more complex RP datasets. We detail

these first before applying model averaging across different models used on each dataset.

3.1.1 UK data

The first dataset we use relies on a SC dataset where public transport commuters living in the UK each make ten choices between three routes. A total of 368 participants completed the survey resulting in 3,680 choices. Each choice task involves an invariant reference trip and two hypothetical alternatives. Each alternative is described by travel time (in minutes), fare (in £), rate of crowded trips, rate of delays (both out of 10 trips), the average length of delays (across delayed trips) and the cost and availability of a delay information service (in £). This dataset has previously been used for decision rule heterogeneity (Hess and Stathopoulos, 2013) as well as for ANA work (Hess et al., 2013b), making it an ideal case study for contrasting model averaging against latent class models.

3.1.2 Sydney dataset

The second dataset that we use for model averaging comes from a Household Travel Survey (HTS-06) that was carried out in Sydney between 2004 and 2006 (Bureau of Transport Statistics, 2012). For this dataset, seven possible modes are established (car driver, car passenger, taxi, walk, bicycle, bus or train) and a large number of destination zones are defined (2,277 travel zones). For the purposes of this paper we consider only 5,173 home-work tours. Level of service and attraction measures were assembled such that attributes could be derived for travel times, costs, waiting times and distances. For a full description of the data and its components, readers should refer to Fox (2015).

3.1.3 California dataset

The final dataset comes from the 2012 California Household Travel Survey (California Department of Transportation, 2013). For this dataset, there are 6,718 choices, with car, bus, rail and air as mode alternatives and 58 destination zones (the different counties in California). Again, we have attraction attributes for the different destinations and times, costs, and distances associated with the different travel modes.

3.2 Model averaging in estimation

Our first aim is to use model averaging to improve model fit. We can also test whether these models are overfitting through the use of validation subsets. Whilst there are many examples of possible uses of model averaging for travel

behaviour modelling, we consider three different datasets to cover three common issues: the specification of heterogeneity, the definition of non-linearity, and the nesting structure in models for large-scale datasets.

3.2.1 Model averaging for specification of heterogeneity

A good first example of when model averaging might be useful is in the consideration of distributions for parameters within a mixed logit (MMNL) model. There is extensive literature on the choice of distributions and it is often clear that different specifications yield relatively similar fit but often substantially different model outputs, making the choice of a final distribution difficult for analysts (Börjesson et al., 2012; Hess et al., 2017), while the use of non-parametric distributions is still beyond the reach of most modellers despite recent innovations on this approach (Fosgerau and Mabit, 2013).

For the UK dataset, we first test the use of continuous distributions. For fare, time, crowding and rate of delays, we use either negative log-normal or negative log-uniform distributions depending on the model (see Table 7.1). We use negative log-normal distributions for the remaining four attributes. This results in 16 different MMNL models, for which the model fits are given in Table 7.1, where we also show the percentage of individuals whose choices are best described by each model (labelled as 'best fit' in Table 7.1).

The best fitting model here is version 15, which has negative log-uniform distributions for fare, time and crowding. A model with negative log-normal distributions for all parameters actually has better fit for more individual participants (13.59% compared to 7.07%) and consequently this model must have a larger range of fits for the individuals to have worse overall fit, which could be a result of long tails (Hess et al., 2017). However, more crucially, there is not much difference between the model fits and this means that there is scope for model averaging.

We apply model averaging across the 16 mixed logit models, i.e. estimating the 16 model specific weights[4], where we do not make these individual specific in our application. This results in a log-likelihood of -2,945, which as expected is better than that of any of the individual models. No formal statistical test is used here as it is not a process of simultaneously estimating all the parameters for all the models on a single dataset. The estimated weights are given in the 'MA share' column in Table 7.1. We see that the model with the best individual log-likeliood obtains the largest share but in addition see non-trivial shares for a substantial subset of other models. Crucially, this includes model 1, which had the worst individual fit, but also the largest share of respondents where this model produced the best fit out of all 16 models. This confirms that model averaging can be a successful approach for incorporating results from models that work well for only a subset of individuals.

---

[4]We use a logit model for class allocation, with 15 constants estimated.

Table 7.1: Log-likelihoods for 16 MMNLs with different combinations of distributions for the UK dataset

MMNL	Time	Fare	Crowding	Rate of delays	Log-likelihood	Best fit	MA Share
1	lognormal	lognormal	lognormal	lognormal	-3,034.16	**13.59%**	7.18%
2	lognormal	lognormal	lognormal	loguniform	-3,030.67	5.16%	0.00%
3	lognormal	lognormal	loguniform	lognormal	-3,019.60	4.62%	0.00%
4	lognormal	lognormal	loguniform	loguniform	-3,015.35	4.35%	0.00%
5	lognormal	loguniform	lognormal	lognormal	-3,027.83	7.34%	0.00%
6	lognormal	loguniform	lognormal	loguniform	-3,015.46	8.42%	7.84%
7	lognormal	loguniform	loguniform	lognormal	-3,001.06	3.80%	0.00%
8	lognormal	loguniform	loguniform	loguniform	-2,996.96	4.35%	3.26%
9	loguniform	lognormal	lognormal	lognormal	-2,982.40	6.79%	3.73%
10	loguniform	lognormal	lognormal	loguniform	-2,983.74	8.15%	15.21%
11	loguniform	lognormal	loguniform	lognormal	-2,980.24	5.43%	14.65%
12	loguniform	lognormal	loguniform	loguniform	-2,990.15	6.25%	0.00%
13	loguniform	loguniform	lognormal	lognormal	-2,982.85	4.08%	0.00%
14	loguniform	loguniform	lognormal	loguniform	-2,978.60	5.43%	9.70%
15	loguniform	loguniform	loguniform	lognormal	**-2,963.14**	7.07%	**34.70%**
16	loguniform	loguniform	loguniform	loguniform	-2,985.48	5.16%	3.73%

We also test to see whether the results from model averaging are overfitting
by using out-of-sample validation. In this case, we split the dataset into five
sections. For each section, we first estimate the parameters for all 16 mixed
logit models individually on the first 80% of the data before calculating the
log-likelihood of the remaining 20% validation set with the estimated param-
eters found for the initial set. We then apply model averaging across the 16
MMNLs for each of the five estimation subsets, before applying the resulting
model averaging structure in each set to the appropriate holdout sample. The
results of this are also shown in Table 7.2, where, for space reasons, we only
ever show the fits for the five most contributing MMNLs in model averaging.

Across both the estimated and forecasted subsets, we consistently see
that the model averaging approach has better fit than that of the best fitting
MMNL model for each estimation subset. Note that across the five different
subsets, four different combinations of distributions result in the best model
fit (Models 14, 12, 15, 9 and 15 respectively across the different subsets). This
highlights the difficult task of choosing distributions and further reinforces
the potential benefits of model averaging. In addition, the MMNL model
that offers the best performance in estimation is not the one with the best
performance in the holdout sample in three out of five cases, while model
averaging always produces a log-likelihood on the holdout that is at least as
good as the best MMNL fit. As in the full sample, we again see that models
that do not fit well across the subset can still contribute to the model average,
with the best fitting model only twice receiving the largest share across the
five subsets, and 13 out of the 16 models are at least once one of the top
five contributors to the model average. Additionally, no single model is the
largest contributor to more than one holdout subset.

3.2.2 Model averaging for linearity assumptions

We next test model averaging on revealed preference (RP) datasets, which
can be more complex. As choices in RP data often include both mode and
destination choice, the models that can be applied often have to be simpler
due to both the vast size of some RP datasets and also due to the large
number of alternatives generated if a modeller is trying to predict the precise
zone or area an individual has chosen to travel to, together with the mode.
Consequently large-scale models are often simple in structure as our usual
more complex models such as mixed logit quickly become computationally
infeasible. Model averaging avoids computational problems by creating a
more complex model from averaging across a number of simpler models.

A key interest in large scale modelling is the specification of the utility
function notably in terms of linearity assumptions (Daly, 2010; Stathopoulos
and Hess, 2012). We therefore trial combinations of parameters for models
on our Sydney HTS-06 data. In these models we use just the mode choice,

Table 7.2: Estimation and holdout sample results for model averaging for the UK dataset

	Best individual MMNL version	Best individual MMNL LL	Estimation					Holdout Sample			
			Model averaging LL	Most contributing MMNLs Version	LL	Share	MA LL Improvement	Model averaging LL	Individual MMNLs Version	LL	MA LL Improvement
Full	15	-2,963	-2,945	15	-2,963	34.7%	18	n/a			
				10	-2,984	15.2%	39				
				11	-2,980	14.7%	35				
				14	-2,963	9.7%	18				
				6	-3,015	7.8%	70				
Holdout 1	14	-2,347	-2,327	11	-2,355	20.7%	28	-625	11	-632	7
				10	-2,350	18.6%	23		10	-637	12
				13	-2,354	14.5%	27		13	-629	4
				2	-2,390	10.7%	63		2	-653	28
				14	-2,347	9.8%	20		14	-629	4
Holdout 2	12	-2,405	-2,383	12	-2,405	24.7%	22	-558	12	-562	4
				9	-2,422	19.2%	39		9	-565	7
				16	-2,408	18.5%	25		16	-564	6
				6	-2,424	14.4%	41		6	-573	15
				3	-2,438	11.1%	55		3	-572	14
Holdout 3	15	-2,354	-2,326	16	-2,356	17.6%	30	-622	16	-626	4
				8	-2,356	15.0%	30		8	-631	9
				15	-2,354	13.9%	28		15	-629	7
				13	-2,369	12.5%	43		13	-627	5
				1	-2,413	12.1%	87		1	-633	11
Holdout 4	9	-2,362	-2,333	8	-2,362	24.1%	29	-615	8	-622	7
				9	-2,362	20.4%	29		9	-635	20
				3	-2,377	18.9%	44		3	-628	13
				15	-2,371	10.3%	38		15	-615	0
				12	-2,370	8.5%	37		12	-629	14
Holdout 5	15	-2,378	-2,347	15	-2,378	22.7%	31	-587	15	-595	8
				6	-2,388	22.5%	41		6	-597	10
				12	-2,396	10.8%	49		12	-601	14
				9	-2,392	8.7%	45		9	-593	6
				11	-2,381	7.3%	34		11	-597	10

for which there are a total of 5,173 choices, each with 7 alternatives. As
there are a number of level of service attributes across the alternatives, we
use four main attribute types and trial each with or without a logarithmic
transformation applied to the set of attributes. The four parameter types
that we consider are costs sensitivities (three different income groups), in-
vehicle travel time sensitivities (bus, car, train, bus connection for train),
other times sensitivities (access time, time until next service, time until sub-
sequent service) and distance sensitivities (car, walking and bus distances).
We additionally have a number of socio-demographic measures included in
the specification of the models, which are based on a model for both mode
and destination (detailed in Table 4.11 of Fox 2015). As we do not consider
destination choice here, we do not use attraction variables. We trial all 16 dif-
ferent combinations of linear and logarithmic transformations of attributes.
This gives us the model results displayed in Table 7.3.

Table 7.3: Results from combinations of linear and logarithmic transformations of at-
tributes on the Sydney HTS-06 mode choice data

Model	Cost	IVT	OT	Distance	Best fit	MA16 Share	Log-likelihood
1	linear	linear	linear	linear	5.2%	0.0%	-2,784.74
2	linear	linear	linear	log	5.5%	0.0%	-2,803.43
3	linear	linear	log	linear	6.5%	66.4%	-2,771.52
4	linear	linear	log	log	10.0%	0.0%	-2,792.17
5	linear	log	linear	linear	4.3%	0.0%	-2,806.83
6	linear	log	linear	log	4.7%	6.6%	-2,814.47
7	linear	log	log	linear	3.3%	0.0%	-2,800.51
8	linear	log	log	log	8.4%	0.0%	-2,804.25
9	log	linear	linear	linear	4.1%	0.0%	-2,801.99
10	log	linear	linear	log	1.6%	0.0%	-2,799.90
11	log	linear	log	linear	5.5%	0.0%	-2,791.18
12	log	linear	log	log	2.9%	7.7%	-2,792.10
13	log	log	linear	linear	6.1%	0.0%	-2,839.87
14	log	log	linear	log	6.4%	19.4%	-2,823.12
15	log	log	log	linear	5.6%	0.0%	-2,838.38
16	log	log	log	log	8.6%	0.0%	-2,818.69
Model averaging across 16 models							-2,750.49

The best performing individual model (model 3) comprises of linear costs,
in-vehicle travel times and distances but a logarithmic transformation for
other travel times. When applying model averaging across the 16 simpler
models, this model obtains 66% of the allocation. Crucially, the improvement
from model averaging across the simpler models is 21 log-likelihood units.
Notably, the second largest share goes to model 14, which is an opposite to
model 3, in that it has a logarithmic transformation for cost, in-vehicle travel
times and distances but not for other travel times. The results suggest that
model 3 provides the best fit due to it providing a steady performance for
each observation. Model 4, as a contrast, is the 2nd best fitting model for
the largest number of choices, but overall performs worse than model 3 by 20
units, demonstrating that it predicts some choices very well and others very

badly. Consequently, the joint model established from model averaging is far less sensitive to outliers, which only have a strong impact if they are not well described by any of the contributing models.

Again, we trial model averaging across models run on the full dataset as well as models run on 80% estimation subsets and 20% validation subsets (See Table 7.4). Across all five holdout samples, model 3 again performs best

Table 7.4: Model averaging log-likelihoods across the attribute treatment combinations for estimation and holdout samples for the Sydney HTS-06 mode choice data

	Full	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Best individual	-2,771.52	-2,230.51	-2,162.17	-2,215.39	-2,229.62	-2,230.61
Model averaging	-2,750.49	-2,216.40	-2,145.19	-2,198.72	-2,212.30	-2,208.86
Best individual (from est.)	n/a	-544.80	-614.94	-561.49	-548.73	-553.30
Model averaging	n/a	-538.86	-610.92	-556.53	-546.15	-549.31

in estimation. This is very different from the case of the mixed logit examples discussed earlier. However, in line with previous results, we again find that estimation and holdout model fits are consistently improved by averaging across all 16 models.

3.2.3 Model averaging for nesting structure

We use our California dataset to test model averaging across models estimating both mode and destination choice. We start by using a multinomial logit model (MNL), two nested logit models (mode over destination, NL (M>D), and destination over mode, NL (D>M)) and a cross nested logit model (CNL). A full description of these models is given by Outwater et al. (2015). We then apply model averaging over these four models. Moving to increasingly complex models results in improvements in model fit and this pattern continues as we move to a model averaging approach (MA), which results in a substantial improvement in model fit over the cross nested logit model (see Table 7.5).

Table 7.5: The results from model averaging (MA) across four basic models applied to the California dataset

Model	MNL	NL (D>M)	NL (M>D)	CNL	MA
Log-likelihood	-19,276	-19,271	-19,220	-19,152	-19,099
Improvement over MNL		5	56	124	177
Improvement over NL(D>M)			51	119	172
Improvement over NL(M>D)				68	121
Improvement over CNL					53
Model averaging share	0.01%	0.02%	35.57%	64.40%	
Proportion of best fits for individuals	5.59%	19.70%	20.66%	54.05%	

Whilst the majority of the model averaging share is given to cross nested logit, model averaging improves model fit by a further 53 log-likelihood units by also giving a substantial share to the nested logit model with mode over

destination. The implications of using a combined model with CNL and this nested logit model are further investigated through consideration of averaging across the model outputs in terms of the differences in elasticities (see Section 3.3.2) and estimates for the value of travel time (see Section 3.4.3).

3.3   Elasticities from model averaging

Elasticities are a key output from choice models estimated on RP data. We now look at the implications of model averaging in this context. Given that elasticities from different models can be very contrasting, a further use of model averaging is that it can be used to derive an 'average' single elasticity. A number of elasticities from different models can be used, with an appropriate weight for the relative importance/performance of the model included. We test this using our two revealed preference datasets.

3.3.1   Sydney elasticities

Our first test of elasticities from model averaging uses the Sydney mode only choice data. This is a particularly relevant example, as elasticities from models with a logarithmic transform for cost are often too low, whilst linear cost models are often too high (Fox et al., 2009). The elasticities for a 10% increase in car costs are shown for the 17 different models tested in Table 7.6. It is noticeable that, whilst many of the elasticities across the different models are similar, the values estimated for train, bus and walking vary more substantially. Unsurprisingly, models with a logarithmic transformation of costs (models 9-16) tend to estimate lower values for alternatives that cost money and higher values for alternatives that do not have an associated cost. This is particularly the case for train and bus, for which elasticity values estimated by models 9-16 are up to half those of values estimated by models 1-8. It is worth noting that as only mode choice is estimated here, the car elasticities observed are lower than those typically observed (see Fox (2015) for elasticities from models predicting mode and destination choice for this data). Whilst model averaging gives a larger share to linear cost models, lower elasticities for train and bus (relative to model 3, which is likely to have been used if outputs from a single model were to be chosen) are found for the model average. As a result, it appears that model averaging may be able to avoid the issues of finding elasticities that are either too high or too low.

3.3.2   California elasticities

We also test different elasticities for the California dataset, where we estimate car cost and time elasticities for trips, trip length and distance. For number

Table 7.6: Elasticities for a 10% increase in car cost for the Sydney mode choice models.

Model	MA Share 16	Log-likelihood	Elasticities						
			Car Driver	Car Passenger	Train	Bus	Bike	Walk	Taxi
1	0.0%	-2,784.74	-0.11	0.13	0.29	0.24	0.15	0.05	0.15
2	0.0%	-2,803.43	-0.10	0.12	0.26	0.22	0.18	0.05	0.13
3	**66.4%**	**-2,771.52**	**-0.11**	**0.13**	**0.28**	**0.23**	**0.14**	**0.05**	**0.14**
4	0.0%	-2,792.17	-0.10	0.11	0.26	0.21	0.17	0.05	0.12
5	0.0%	-2,806.83	-0.14	0.15	0.35	0.30	0.18	0.06	0.16
6	**6.6%**	**-2,814.47**	**-0.12**	**0.14**	**0.29**	**0.25**	**0.17**	**0.06**	**0.14**
7	0.0%	-2,800.51	-0.14	0.15	0.35	0.29	0.17	0.06	0.16
8	0.0%	-2,804.25	-0.11	0.13	0.29	0.24	0.16	0.05	0.13
9	0.0%	-2,801.99	-0.05	0.09	0.09	0.11	0.14	0.10	0.13
10	0.0%	-2,799.90	-0.08	0.13	0.14	0.16	0.20	0.13	0.18
11	0.0%	-2,791.18	-0.05	0.08	0.09	0.10	0.12	0.09	0.12
12	**7.7%**	**-2,792.10**	**-0.07**	**0.12**	**0.13**	**0.15**	**0.19**	**0.12**	**0.17**
13	0.0%	-2,839.87	-0.06	0.11	0.11	0.13	0.16	0.11	0.15
14	**19.4%**	**-2,823.12**	**-0.08**	**0.13**	**0.14**	**0.17**	**0.20**	**0.13**	**0.18**
15	0.0%	-2,838.38	-0.06	0.09	0.10	0.11	0.14	0.10	0.13
16	0.0%	-2,818.69	-0.07	0.12	0.13	0.15	0.18	0.12	0.16
Model Averaging 16		**-2,750.49**	**-0.10**	**0.13**	**0.24**	**0.21**	**0.16**	**0.07**	**0.15**

of tours, average tour length and total distance respectively, we define:

$$TourElasticity = log(\frac{ForecastedTours}{BaseTours})/log(1.1), \qquad (7.9)$$

$$TourLengthElasticity = log(\frac{ForecastedTourLength}{BaseTourLength})/log(1.1), \quad (7.10)$$

$$DistanceElasticity = log(\frac{ForecastedTotalDistance}{BaseTotalDistance})/log(1.1). \quad (7.11)$$

Elasticities from the two best fitting candidate models (CNL and NL mode over destination, see Section 3.2.3) produce some very different[5] values (see the purple and green bars respectively in Figure 3). Consequently, if a single output for each elasticity is required, model averaging provides suitable values which take into account the relative performance of the different models.

Fig. 7.1: Elasticities from the four different candidate models and model averaging for the California dataset

---
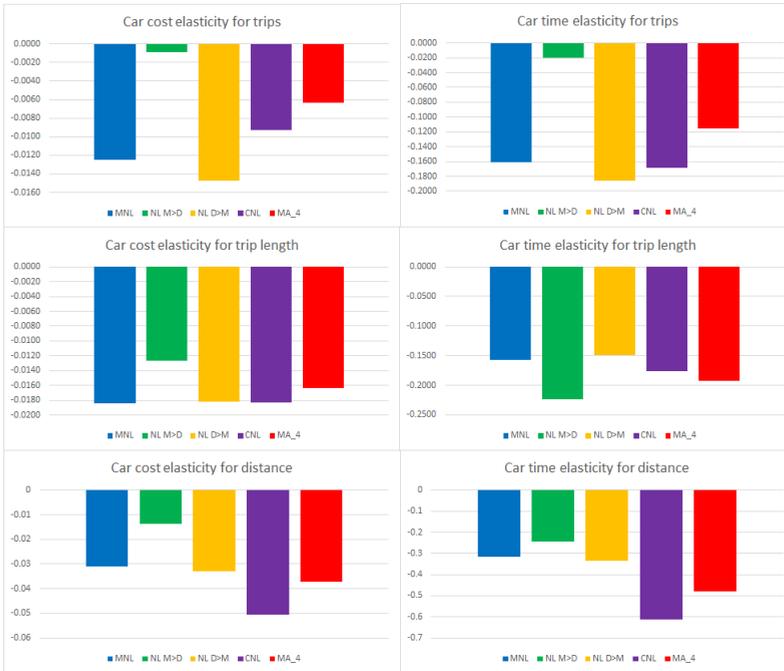
[5]For a full review of the elasticities from these four candidate models, readers should refer to Outwater et al. (2015).

3.4 Willingness-to-pay outputs from model averaging

In this section, we explore welfare outputs from model averaging across the different models for the different datasets. We demonstrate how this can be done for both SP and RP datasets. For our SP dataset, we look at value of travel time as well as values for decreasing both the amount of crowding and the rate of delays. For our RP data, we take a more detailed look at the different estimates for the value of travel time.

3.4.1 Outputs from UK models

For the UK models, we first use the estimates from each of the 16 mixed logit models to obtain values[6] for the value of travel time (VTT, £/hour), value of crowding (VCR, amount paid in £ for 1/10 less crowded trips) and value of the rate of delays (VDE, amount paid in £ for 1/10 less delayed trips). We use the full distributions from the individual models in model averaging - the resulting means and standard deviations of these measures for each model and the model average is given in Table 7.7.

In comparison to the estimates obtained if we had simply used the best fitting mixed logit model (MMNL-15), results from model averaging suggest that the willingness to pay for changes in travel time and the rate of delays are not as high. The opposite is true for changes in the number of crowded trips, for which model averaging produces a higher estimate than MMNL-15. Notably, model averaging predicts a much wider standard deviation for the value of crowding.

3.4.2 Sydney VTT

Given that we use several different mode-specific travel time coefficients and three different income groups for our Sydney models, we can study a number of different travel time outputs from model averaging. We can compare the values for different groups of individuals as we have three cost coefficients in each model for three different income categories (1st: < $26k AUD, 2nd: $26-36.4k AUD, 3rd: > $36.4k AUD). We first obtain the value of travel time from all of the candidate models. As some of the models use logarithmic transformations for costs and times, we multiply these measures by a representative cost ($5.48) and divide by a representative time (49 minutes), as required. These outputs are detailed in Table 7.8.

It appears that, whilst the different models have fairly similar model fit, the value of travel times vary significantly, both across models and modes. The effect of income, however, is fairly consistent, with individuals of a higher income prepared to pay more to reduce time spent travelling. The difference

---

[6]Note that as we use a logarithmic transformation for the cost attribute, we multiply values by 3, as this is the average cost of chosen alternatives (to the nearest pound).

Table 7.7: Welfare measures obtained from the UK models

MMNL	TT	LF	CR	DE	Model LL	VTT mean	VTT sd	VCR mean	VCR sd	VDE mean	VDE sd	model share π_m
1	n	n	n	n	-3.034	3.390	6.812	0.291	0.725	0.265	0.926	7.2%
2	n	n	n	u	-3.031	3.332	5.866	0.321	0.921	0.244	0.723	0.0%
3	n	n	n	n	-3.020	3.099	5.538	0.381	1.275	0.231	0.675	0.0%
4	n	n	n	n	-3.015	3.382	6.532	0.396	1.367	0.274	0.820	0.0%
5	n	n	n	u	-3.028	3.195	4.755	0.366	1.252	0.258	0.868	0.0%
6	n	n	n	u	-3.015	3.113	4.399	0.441	2.866	0.228	0.438	7.8%
7	n	n	u	n	-3.001	3.672	6.221	0.375	0.813	0.228	1.052	0.0%
8	n	n	u	u	-2.997	2.945	4.226	0.319	0.695	0.185	0.468	3.3%
9	n	n	u	u	-2.982	3.809	7.620	0.297	0.803	0.205	0.506	3.7%
10	n	u	n	u	-2.984	3.945	8.084	0.338	0.964	0.231	0.593	15.2%
11	n	u	n	n	-2.980	3.871	8.407	0.405	1.376	0.228	0.625	14.6%
12	n	u	u	n	-2.990	3.962	9.251	0.330	0.969	0.265	0.772	0.0%
13	n	u	u	n	-2.983	3.624	5.730	0.305	0.882	0.256	1.076	0.0%
14	u	u	n	n	-2.979	3.612	5.889	0.310	0.876	0.193	0.348	9.7%
15	**u**	**u**	**u**	**n**	**-2.963**	**3.891**	**6.615**	**0.302**	**0.628**	**0.260**	**0.794**	**34.7%**
16	u	u	u	u	-2.985	3.691	6.406	0.325	0.659	0.199	0.433	3.7%
Model Averaging					**-2.945**	**3.737**	**6.951**	**0.335**	**1.019**	**0.235**	**0.658**	

Table 7.8: Value of travel times (AUD/hr) obtained from the models for the Sydney choice-only data, across different modes and income categories.

Model	LL	Share	Car			Train			Bus			Access		
			1	2	3	1	2	3	1	2	3	1	2	3
1	-2,784.74	0.0%	9.10	12.33	15.33	2.42	3.28	4.08	7.20	9.75	12.12	3.50	4.74	5.89
2	-2,803.43	0.0%	7.77	10.98	13.93	0.12	0.17	0.22	6.15	8.68	11.02	3.40	4.80	6.10
3	**-2,771.52**	**66.4%**	**11.37**	**15.29**	**19.57**	**4.15**	**5.58**	**7.15**	**8.31**	**11.18**	**14.30**	**7.07**	**9.51**	**12.17**
4	-2,792.17	0.0%	9.68	13.51	17.65	1.70	2.37	3.10	7.05	9.84	12.86	7.08	9.88	12.91
5	-2,806.83	0.0%	1.02	1.31	1.56	4.99	6.39	7.61	0.25	0.32	0.38	1.86	2.39	2.84
6	-2,814.47	6.6%	4.37	5.88	7.16	4.46	6.01	7.31	2.12	2.85	3.47	2.77	3.73	4.54
7	-2,800.51	0.0%	2.59	3.32	3.96	4.70	6.03	7.20	0.64	0.82	0.98	5.35	6.85	8.19
8	-2,804.25	0.0%	6.88	9.20	11.34	3.84	5.14	6.33	2.99	4.01	4.94	6.86	9.18	11.31
9	-2,801.99	0.0%	20.88	21.92	21.90	3.34	3.51	3.51	13.68	14.36	14.35	8.15	8.55	8.55
10	-2,799.90	0.0%	11.46	11.91	12.04	0.31	0.33	0.33	7.91	8.22	8.30	4.80	4.98	5.04
11	-2,791.18	0.0%	27.14	28.21	28.40	7.46	7.75	7.80	17.21	17.89	18.00	15.81	16.43	16.54
12	-2,792.10	7.7%	14.35	14.80	15.06	2.52	2.60	2.64	9.58	9.88	10.05	9.53	9.83	10.00
13	-2,839.87	0.0%	3.80	3.94	3.92	11.81	12.25	12.19	-0.57	-0.59	-0.58	4.65	4.82	4.80
14	-2,823.12	19.4%	6.49	6.68	6.71	6.79	6.99	7.03	1.98	2.04	2.05	4.04	4.16	4.18
15	-2,838.38	0.0%	8.38	8.59	8.60	12.32	12.64	12.65	0.88	0.91	0.91	12.65	12.97	12.99
16	-2,818.69	0.0%	10.90	11.11	11.24	6.33	6.45	6.53	3.74	3.81	3.85	10.11	10.30	10.43
MA	**-2,750.49**		**10.19**	**12.96**	**15.91**	**4.56**	**5.66**	**6.79**	**6.77**	**8.76**	**10.89**	**6.39**	**8.12**	**9.95**

235

between models is very significant, with the results from some models suggesting that individuals are willing to spend up to 10 times more than other models suggest. This means that if we were to pick a single model to use the outputs from, very different interpretations of the value of travel time could be made depending on which model is chosen. Without model averaging, it may be hard to know which models to rely more heavily on. The results from model averaging, however, appear reasonable.

3.4.3 California VTT

For our California data, we can calculate four mode-specific values of travel time from each of the different models. The results of these models are given in Table 7.9. In this case, as there are only two models that contribute to

Table 7.9: Value of travel time estimates by mode across the different models for the California data

		MNL	NL (D>M)	NL (M>D)	CNL	**MA**
Log-likelihood		-19,276	-19,271	-19,220	-19,152	**-19,099**
VTT	car	73.10	71.70	135.94	88.09	**105.11**
	bus	78.75	77.66	141.72	98.33	**113.76**
	rail	72.15	73.15	67.41	77.68	**74.03**
	air	24.01	25.85	3.84	42.40	**28.68**
Model share		0.01%	0.02%	35.57%	64.40%	

the model average, model averaging provides a value that is close to halfway between the estimates for the value of travel time from the CNL model and the nested logit model with mode over destination. For air, this results in an estimate that is actually closer to the non-contributing MNL value.

3.5 Information processing work

We first look at the case of ANA, where we adopt a specification in line with Hess et al. (2013b).

  We first estimate a simple MNL model, where we use a logarithmic transform on the fare attribute given earlier evidence of strong non-linearity. This model uses five marginal utility parameters for the continuous attributes, two parameters for the dummy coded delay information system, and two alternative specific constants (ASC). The results for this model are shown in Table 7.10 where all estimates are of the correct sign.

  We next move to the latent class model for attribute non-attendance. We use a model with 2^K classes, with all combinations of attendance and non-attendance for the K parameters. The probability for class s is given by

Table 7.10: MNL results for public transport route choice

LL(final)	-3,366.95	
ρ^2	0.1672	
adj. ρ^2	0.165	
	Estimate	Rob.t.ratio(0)
ASC_1	0.3841	5.76
ASC_2	0.1608	3.26
β_{tt}	-0.0467	-9.47
$\beta_{\text{log-fare}}$	-5.9726	-18.89
β_{crowding}	-0.2198	-8.51
$\beta_{\text{rate of delays}}$	-0.2411	-9.82
$\beta_{\text{average delay}}$	-0.0421	-5.35
$\beta_{\text{info system charged}}$	-0.0833	-1.04
$\beta_{\text{info system free}}$	0.3370	5.06

π_s, with $0 \leq \pi_s \leq 1$ and $\sum_{s=1}^{S} \pi_s = 1$. Rather than imposing constraints in estimation, an easier approach is to use $\pi_s = \frac{e^{\delta_s}}{\sum_{m=1}^{S} e^{\delta_m}}$, with one δ_m, i.e. the parameter used in the class allocation probabilities, being fixed to zero. Nevertheless, this specification still involves estimating $2^K - 1$ separate δ terms, of which many will be very negative, equating to very small class probabilities. In the context of the applications presented in this paper, we make use of a simplified approach, by instead setting

$$\pi_s = \prod_{k=1}^{K} \left(\Lambda_{s,k} \left(1 - P_{\text{N-A},k} \right) + \left(1 - \Lambda_{s,k} \right) P_{\text{N-A},k} \right), \qquad (7.12)$$

where $\Lambda_{s,k}$ gives the entry in Λ relating to attribute k in class s, where this is 1 only if attribute k is attended to in class s. With this specification, we only need to estimated K separate δ elements (with $P_{\text{N-A},k} = \frac{e^{\delta_k}}{e^{\delta_k}+1}$), as opposed to $2^K - 1$, leading to significant reductions in the number of parameters.

The results for this model are shown in Table 7.11. We see an improvement in log-likelihood by 308.16 units for 7 additional parameters. This is highly significant and in line with previous findings when using such a confirmatory latent class model for ANA. We also see that the parameters in the attendance classes have increased substantially, where this is in line with the notion that the MNL model would find an intermediary value between 0 for the non-attenders and a positive value for those attending to the attribute. However, the implied rates of non-attendance are unrealistically high, exceeding 50% for all attributes except fare.

We finally look at the estimation of our model averaging structure. For this, we first estimate 128 individual models, corresponding to all possible

Table 7.11: Confirmatory latent class model for attribute non-attendance

LL(final)	-3,058.79	
ρ^2	0.2434	
adj. ρ^2	0.2395	
	Estimate	Rob.t.ratio(0)
ASC_1	0.8416	10.32
ASC_2	0.329	4.23
β_{tt}	-0.1841	-5.64
$\beta_{\text{log-fare}}$	-14.6889	-14.37
β_{crowding}	-1.1524	-7.16
$\beta_{\text{rate of delays}}$	-1.1307	-5.62
$\beta_{\text{average delay}}$	-0.3966	-4.85
$\beta_{\text{info system charged}}$	2.3264	3.37
$\beta_{\text{info system free}}$	2.0433	7.23
$\delta_{\text{NA,tt}}$	0.3232	1.11
$\delta_{\text{NA,log-fare}}$	-0.5142	-3.43
$\delta_{\text{NA,crowding}}$	0.7767	3.3
$\delta_{\text{NA,rate of delays}}$	0.7363	2.43
$\delta_{\text{NA,average delay}}$	1.1917	4.02
$\delta_{\text{NA,info system charged}}$	3.1776	3.82
$\delta_{\text{NA,info system free}}$	0.9874	3.61
	Implied rate of NA	
	Estimate	Rob.t.ratio(0)
travel time	0.5801	8.18
fare	0.3742	10.65
crowding	0.685	13.49
rate of delays	0.6762	10.21
average delay	0.767	14.48
info system charged	0.96	30.05
info system free	0.7286	13.47

combinations of attribute attendance and non-attendance, i.e. going from a model with all 9 model parameters to one with the two ASCs only. We then estimate the model averaging structure, where we again use multiplicative class allocation probabilities, as in the LC model. We initially estimate seven class allocation weights as in the LC model but find that four the first four attributes, the constants go towards $-\infty$, suggesting a zero probability of ANA.

The results of the model averaging work are shown in Table 7.12. We see that this model now only offers a marginally better log-likelihood than

Table 7.12: Model averaging for ANA work

LL(final)	-3,363.28			Implied rate of NA		
	Estimate	Rob.t.ratio(0)			Estimate	Rob.t.ratio(0)
$\delta_{NA,\text{average delay}}$	1.1917	4.02		average delay	0.129	1.17
$\delta_{NA,\text{ch inf sys}}$	3.1776	3.82		info system charged	0.5211	1.22
$\delta_{NA,\text{free inf sys}}$	0.9874	3.61		info system free	0.2399	2.33

Information for 8 retained models								
individual LL	-3,367.75	-3,366.95	-3,400.98	-3,390.17	-3,391.85	-3,391.62	-3,424.48	-3,416.22
ranking	2	1	6	3	5	4	8	7
best fitting N	12	17	14	14	9	8	12	9
MA share	34.50%	31.71%	10.89%	10.01%	5.11%	4.70%	1.61%	1.48%
attribute included								
travel time	YES	YES	YES	YES	YES	YES	YES	YES
fare	YES	YES	YES	YES	YES	YES	YES	YES
crowding	YES	YES	YES	YES	YES	YES	YES	YES
rate of delays	YES	YES	YES	YES	YES	YES	YES	YES
average delay	YES	YES	YES	YES	NO	NO	NO	NO
info system charged	NO	YES	NO	YES	NO	YES	NO	YES
info system free	YES	YES	NO	NO	YES	YES	NO	NO
est. (rob. t-rat)								
ASC_1	0.41 (6.46)	0.38 (5.76)	0.4 (6.24)	0.32 (4.77)	0.39 (6.15)	0.38 (5.61)	0.38 (5.91)	0.31 (4.61)
ASC_2	0.16 (3.29)	0.16 (3.26)	0.16 (3.29)	0.16 (3.17)	0.18 (3.59)	0.18 (3.58)	0.17 (3.46)	0.17 (3.41)
β_{tt}	-0.05 (-9.48)	-0.05 (-9.47)	-0.05 (-9.73)	-0.05 (-9.63)	-0.05 (-9.35)	-0.05 (-9.34)	-0.05 (-9.59)	-0.05 (-9.49)
$\beta_{\text{log-fare}}$	-5.95 (-18.86)	-5.97 (-18.89)	-5.77 (-18.19)	-5.9 (-18.62)	-5.87 (-18.81)	-5.88 (-18.81)	-5.68 (-18.12)	-5.8 (-18.51)
β_{crowding}	-0.22 (-8.5)	-0.22 (-8.51)	-0.22 (-8.59)	-0.22 (-8.61)	-0.22 (-8.46)	-0.22 (-8.46)	-0.22 (-8.55)	-0.22 (-8.56)
$\beta_{\text{rate of delays}}$	-0.24 (-9.76)	-0.24 (-9.82)	-0.24 (-9.82)	-0.24 (-9.95)	-0.27 (-10.94)	-0.27 (-10.98)	-0.26 (-11)	-0.27 (-11.17)
$\beta_{\text{average delay}}$	-0.04 (-5.32)	-0.04 (-5.35)	-0.04 (-5.29)	-0.04 (-5.51)	0	0	0	0
$\beta_{\text{ch inf sys}}$	0	-0.08 (-1.04)	0	-0.27 (-3.67)	0	-0.04 (-0.57)	0	-0.24 (-3.24)
$\beta_{\text{free inf sys}}$	0.36 (5.96)	0.34 (5.06)	0	0	0.36 (5.91)	0.35 (5.22)	0	0

the MNL model in Table 7.10, much in contrast with the LC model in Table 7.11. In addition to the earlier finding of zero weight for any classes that imply non-attendance of either time, fare, crowding or the rate of delays, we see low rates for average delay and the free information system, with a higher rate for the charged system. We further see that the 8 models that obtain the best individual log-likelihoods are also the only 8 models that contribute to the model average. However, the best ranking model individually is not necessarily the one contributing the most to the model average. Finally, out of the 368 individuals in the data, only 95 have their choices explained the best way by one of these 8 models, where a remarkable 104 out of the 128 models have at least one individual where they are the best performing model.

Overall, the findings from this analysis are much in contrast with those from the confirmatory latent class model in that very little evidence of ANA is found. In addition, there is very little variation in the remaining parameters across classes. Of course, the counter-argument could be that the model averaging approach cannot retrieve ANA as it is based on individual models that each apply a homogenous approach to all individuals. However, some reassurance can be obtained from the fact that the model averaging results are in line with the findings by Hess et al. (2013b) which find evidence of ANA only for the average delay attribute and for the delay information attribute after allowing for random heterogeneity in their models. It is thus doubtful whether additional insights would be obtained with more flexibility for the individual models, such as by including random heterogeneity.

3.6 Decision rule heterogeneity work

We next turn to decision rule heterogeneity. To maximise the possibility of finding such heterogeneity, we consider five very different decision rules, namely:

Multinomial logit (MNL): We assume that the utility a respondent n obtains from alternative i (out of J alternatives) in choice task t is:

$$V_{int} = U_{int} + \epsilon_{int}, \tag{7.13}$$

where V_{int} and ϵ_{int} are the deterministic and random components of utility respectively. The assumption of a type I extreme value distribution for ϵ_{int} then gives us the usual MNL choice probabilities:

$$P_{MNL,int} = \frac{e^{V_{int}}}{\sum_{j=1}^{J} e^{V_{jnt}}}. \tag{7.14}$$

Random regret minimisation (RRM): We base our random regret minimisation (RRM) model on the updated specification of Chorus (2010).

Thus, the deterministic regret for respondent n for alternative i in choice task t is given by:

$$R_{int} = \delta_{RRM,i} + \sum_{k=1}^{K} \sum_{j \neq i} ln(1 + e^{\beta_k (x_{jntk} - x_{intk})}) \qquad (7.15)$$

with $k = 1, ..., K$ is an index across attributes, β_k is a attribute-specific coefficient for attribute k and $\delta_{RRM,i}$ is an alternative specific constant. With the error component of regret also being given by a type I extreme value distribution, the corresponding RRM probabilities for a respondent n choosing alternative i in choice task t is given by:

$$P_{RRM,int} = \frac{e^{-R_{int}}}{\sum_{j=1}^{J} e^{-R_{jnt}}} \qquad (7.16)$$

Decision field theory (DFT): DFT is a dynamic, stochastic model where the preferences for alternatives update over the course of the decision-making process (Busemeyer and Townsend, 1992). Under decision field theory (DFT), a decision-maker stochastically considers the different attributes of the alternatives over the course of a decision-making process. The DFT model utilised in the empirical tests in this paper is based on the version in Chapter 4, which incorporates attribute-specific scaling factors. For a full description of decision field theory, how it can be applied and how the different parameters in the model work, readers should consult Chapters 2 & 4.

Quantum pairwise comparison (QPCA) Our quantum model is based on the first model (quantum pairwise comparison framework A) defined in Chapter 5. Under a quantum model, the possible choice alternatives can be represented by a set of orthogonal vectors which make up the basis for a multidimensional Hilbert space (Bruza et al., 2015). A decision-maker's opinion or 'state' can then be represented by another vector within this space. The action of making a choice is then represented by a projection from this state vector onto the vector representing the chosen alternative (see Figure 5.2). Allowing the state vector to be of unit length results in the set of squared projection lengths onto each of the possible alternatives summing to one. Under QPCA, the length of projection for alternative i (for respondent n in choice task t) is:

$$|\rho_{int}| = \delta_{QPCA,i} + I_0 + \sum_{k=1}^{K} \sum_{j \neq i} wt_{ij} \cdot ln(1 + e^{\beta_k (x_{intk} - x_{jntk})}), \quad (7.17)$$

where $\delta_{QPCA,i}$ are alternative-specific constants, I_0 is a constant that has the same value across all alternatives, wt_{ij} is a weight for the relative importance of the comparison between alternatives i and j and β_k

is a coefficient for attribute k as before for RRM. Once these projection lengths have been calculated, the probability for each alternative can be defined simply as:

$$P_{QPCA,jnt} = \frac{|\rho_{jnt}|^2}{\sum_{i=1}^{J}(|\rho_{int}|^2)}, \qquad (7.18)$$

where $i = 1,...J$ is an index across the possible alternatives.

Relative advantage maximisation (RAM) In RAM (Leong and Hensher, 2014), the utility for respondent n in choice task t is:

$$U_{int} = \delta_{RAM,i} + \sum_{k=1}^{K} \beta_k \cdot x_{intk} + \sum_{j \neq i} RA(i,j), \qquad (7.19)$$

which is equivalent to a multinomial logit model with the addition of the comparison of relative advantages $RA(i,j)$ of alternative i in comparison to each of the other alternatives. This relative advantage is then defined:

$$RA(i,j) = \frac{A(i,j)}{A(i,j) + D(i,j)}, \qquad (7.20)$$

where the advantages are calculated $A(i,j) = ln(1 + e^{\beta_k(x_{intk} - x_{jntk})})$ and the disadvantages $D(i,j) = ln(1 + e^{\beta_k(x_{jntk} - x_{intk})})$.

For our SP dataset, we first apply the five different models individually, obtaining the results given in Table 7.13. We see that DFT obtains the best log-likelihood ahead of QPCA, with the performance of the three logit-style models is poorer and comparatively more similar. As a first step, we look at model averaging across all five models applied to this dataset, where the resulting shares and fit are shown in Table 7.13. We see that the model average leads to a further small improvement in model fit over the best fitting individual model, i.e. DFT, where this model also obtains by far the largest share in the model average. As with earlier examples, the shares are not necessarily proportional to the model fit of the individual model, and we see that RRM obtains a substantially larger share than QPCA, despite having poorer overall individual log-likelihood. This again shows that some models can work well for some people even if they obtain a lower overall fit to the sample.

In practice, the estimation of a latent class model with five separate classes all using individual decision rules is computationally challenging and most applications rely on just combining a couple of different rules. We therefore look at the estimation of 15 different latent class structures with two classes per model, thus also allowing for five models where the two classes are of the same type, i.e. looking for taste heterogeneity alone. Table 7.14 gives

Table 7.13: Results from different individual models applied to the SP dataset

Model	Type	Log-likelihood	BIC	MA Share
1	MNL	-3,360.43	6,803	0.00%
2	RRM	-3,363.91	6,810	17.67%
3	DFT	-3,317.18	6,749	76.54%
4	QPCA	-3,336.44	6,771	5.70%
5	RAM	-3,354.55	6,791	0.08%
Model averaging		-3,312.40		

the log-likelihoods of these models. For all 15 models, a likelihood ratio test against the corresponding model (in the case of single decision rule) or two corresponding models (in the case of two decision rules) clearly rejects the base model. This would provide evidence of taste heterogeneity (in the case of single structure models) and would typically be seen as evidence of decision rule heterogeneity in the case of the models with two different structures in the two classes.

Most existing applications compare a model combining multiple different decision rules to a set of single class models using the individual rules. This comparison is of course likely to be biased in the presence of taste heterogeneity. Crucially, the improvements to be made from combining different structures depend on their individual performance. For example, we see that, for DFT, which is the best performing individual model in Table 7.13, combining the model with a different structure does not reach as high a log-likelihood as a structure with two separate DFT classes, although a better BIC may be obtained. On the other hand, for those models that perform less well individually, combining them with a different structure gives a better log-likelihood than a model with two classes using the same structure. This already suggests that the results from the latent class structure point more towards taste heterogeneity than decision rule heterogeneity. Further insights are detailed in Table 7.15, which for each pair of different decision rules (x, y), gives the difference in model fit between this model and the better fitting model from the latent class models with x in both classes or y in both classes[7]. We see only two cases in favour of decision-rule heterogeneity. The MNL-RRM model outperforms RRM-RRM by 3.67 log-likelihood units (as well as the MNL-MNL model by 10.47 units). Additionally, QPC-RAM has a better log-likelihood than either QPC-QPC or RAM-RAM. However, all other differences are negative, indicating that models with the same decision rule in the 2 different classes frequently perform just as well or better than models with differing decision rules.

---

[7] Note that no formal fit comparisons are made here.

Table 7.14: Results from latent class models applied to SP dataset

Model	Class 1	Class 2	Log-likelihood	BIC	MA Share
1	MNL	MNL	-3,113.13	6,399	0.0%
2	MNL	RRM	-3,102.66	6,378	0.0%
3	MNL	DFT	-3,099.84	6,380	6.7%
4	MNL	QPC	-3,106.76	6,394	0.0%
5	MNL	RAM	-3,100.79	6,374	0.0%
6	RRM	RRM	-3,106.33	6,385	16.0%
7	RRM	DFT	-3,086.79	6,354	11.7%
8	RRM	QPC	-3,096.35	6,373	0.0%
9	RRM	RAM	-3,104.22	6,381	0.0%
10	DFT	DFT	-3,077.79	6,361	52.8%
11	DFT	QPC	-3,085.28	6,376	0.0%
12	DFT	RAM	-3,085.38	6,351	0.0%
13	QPC	QPC	-3,095.71	6,380	12.8%
14	QPC	RAM	-3,094.59	6,370	0.0%
15	RAM	RAM	-3,100.27	6,373	0.0%
Model Averaging			-3,071.46		

Further evidence is given in the model averaging results in Table 7.14. We see that model averaging obtains a better log-likelihood than any of the individual LC models, in line with the previous results in this paper. Crucially, however, 81.6% of the share is given to models that each time use just a single decision rule, again highlighting the importance of within-model taste heterogeneity, at least for this data.

Table 7.15: Differences in log-likelihood between combinations of rules and best fitting model using same rule in both classes

MNL	3.67	-22.05	-11.06	-0.52
	RRM	-9.00	-0.64	-3.95
		DFT	-7.49	-7.59
			QPC	1.12
				RAM

We explore the best example for decision-rule heterogeneity (MNL-RRM) in more detail by also considering the outputs for the parameter estimates, in comparison to a model average performed on MNL and RRM. The results for this are shown in Table 7.16. For each model we have coefficients for

travel time (TT), log of the fare (LFare[8]), rate of crowding (Crowd), length of delays (Delay), rate of delays (Rate), a reliability level (Rel, created by calculating the expected length of delays), and the provision of a charged delay information service (Inf) or a free service (InfF). Finally, we include two alternative specific constants for the first two alternatives. Table 7.16 gives model fit as well as estimates for the above parameters for both a latent class model and a model averaging approach. The model averaging approach separately runs MNL and RRM models before then estimating a class allocation parameter individually. Crucially, the model averaging approach does not result in a significant improvement over a MNL model on its own, with an improvement of just 0.07 log-likelihood units. As a contrast, the latent class approach results in a vast improvement in model fit (258 units). At face value, this would again suggest decision rule heterogeneity, although the fit is not much better than for the MNL-MNL or RRM-RRM models. Most significantly, it appears that the fare parameter estimates (highlighted in red) are very different between the two classes. In contrast with the model averaging results, and given the poor class specific model fit for the RRM class (compared to the RRM-RRM model), we believe that this finding shows that a substantial share of the improvements obtained by this model are due to heterogeneity in the cost sensitivity rather than heterogeneity in the decision rules. This means that the classes individually have very poor fit (as they cannot explain all individuals) but when combined into a latent class approach, the result is a model with far superior model fit. Together with the poor improvement from model averaging, these results suggest that most of the model improvement is due to taste rather than decision rule heterogeneity.

4 Conclusions

Despite successful results in a number of fields including health, ecology and economics, model averaging has yet to make a transition into mainstream choice modelling. In this paper, we demonstrate that it is very simple to run and that it consistently improves model fit in both estimation and forecasting. Whilst we apply model averaging through the use of sequential latent class models, other methods are possible, with Bayesian methods used for model averaging typical in other disciplines (Raftery et al., 2005; Wang et al., 2004; Wintle et al., 2003). Consequently, future work could compare different model averaging methods. However, we find that model averaging using a simple sequential latent class structure provides many benefits.

We demonstrate that model averaging can be applied across a large num-

---

[8]Note that we use a log transform of the fare rather than the fare itself as a cost damping affect is observed.

Table 7.16: A detailed example of model averaging compared to a simultaneous latent class approach using MNL and RRM

| | Latent Class - 1 model 21 pars, estimated simultaneously | | Model averaging - 3 models 2*10 pars, then 1 for MA | |
	Class 1:MNL	Class 2:RRM	Class 1: MNL	Class 2: RRM
Class LL:	-3,645.30	-4,431.55	-3,360.43	-3,363.91
Log-likelihood		-3,102.66		-3,360.36
Class LL:	-3,645.30	-4,431.55	-3,360.43	-3,363.91
Log-likelihood		-3,102.66		-3,360.36
asc_{alt1}	0.64 (6.42)	0.04 (0.27)	0.39 (5.85)	0.27 (4.17)
asc_{alt2}	0.25 (2.81)	0.20 (1.13)	0.16 (3.3)	0.17 (3.38)
β_{TT}	-0.05 (-6.74)	-0.05 (-6.79)	-0.05 (-9.5)	-0.03 (-9.58)
β_{Lfare}	-3.21 (-6.1)	-11.32 (-7.58)	-6.00 (-18.87)	-4.11 (-17.66)
β_{Crowd}	-0.31 (-7.41)	-0.15 (-2.89)	-0.22 (-8.58)	-0.15 (-8.59)
β_{Delay}	-0.06 (-1.27)	-0.05 (-1.29)	-0.03 (-3.24)	-0.02 (-3.06)
β_{Rate}	-0.34 (-4.82)	-0.09 (-1.76)	-0.19 (-5.96)	-0.12 (-5.82)
β_{Rel}	-0.05 (-3.22)	0.00 (0.06)	-0.06 (-2.64)	-0.04 (-2.71)
β_{Inf}	-0.10 (-0.82)	-0.16 (-1.09)	-0.09 (-1.13)	-0.05 (-0.95)
β_{InfF}	0.54 (5.84)	0.05 (0.47)	0.33 (4.95)	0.22 (4.85)
π_m	59.30% (10.89)	40.70%	87.70% (2.7)	12.30%

ber of candidate models. These models can be very similar, with model averaging proving effective when used across multiple mixed logit models with various different combinations of distributions for the parameters. The models can also be more different, such as in our nesting structures for large scale modelling. With complex models often infeasible to run when there are hundreds or even thousands of alternatives, model averaging provides a simple and efficient method for improving models, with consistent improvements in model fit found when applying it over a number of simple models. Additionally, model averaging is less sensitive to outliers, as unlikely choices only have an impact on the model fit if they are outliers across all models contributing to the model average. This also means that model averaging is very good at making the most of models which are very accurate at describing some choices but less accurate for others. Consequently, the best fitting model may not contribute to a model average.

We show that model averaging always provides model fit at least as good as the best fitting candidate model. We have purposefully not conducted statistical tests for these improvements in fit. Indeed, model averaging should not be seen as a different model which can be compared to individual structures, such as a simultaneous latent class model with different models in each class. Indeed, for model averaging, the process only involves calculating a weighted average of the outputs from individual models and does not involve the reestimation of the parameters from the individual models, where these always come from individual models estimated on the full sample.

Whilst we only ever consider the use of constants for class allocation, more complex structures could easily be adopted. For example, the param-

eterisation of class allocation within model averaging could be performed very simply by using socio-economic attributes. A final key advantage of model averaging is that it is very easy to apply. A modeller does not even require knowledge of the individual models within the classes to apply model averaging. This means that, for example, practitioners could ask multiple researchers to apply models to the same dataset and then average across the models, for which they would only need the underlying log-likelihood contribution for each individual or observation in the dataset. This may go some way to mitigating risk, as well as having a chance of improving the model. Consequently, there are many advantages to be gained by applying model averaging for both applied and theoretical transport behaviour modellers.

This paper also revisits the use of latent class models to capture different behavioural processes such as attribute non-attendance and decision rule heterogeneity. These approaches have been very popular in recent years and have often been shown to produce significant gains in fit over simpler models. We first argue that many such findings may be due to an unfair comparison with models not allowing for any heterogeneity and that the findings may in fact be driven by taste heterogeneity at the level of a fixed model specification rather than the presence of other phenomena.

We contrast the findings obtained from such latent class models with those obtained using model averaging which combines the evidence from a number of separately estimated models. This latter approach of course leads to inferior model fit compared to a simultaneous latent class model but our findings provide some evidence that suggests that these bigger improvements may indeed be in part due to effects other than those that analysts seek to uncover.

In practice, an analyst should of course attempt to simultaneously allow for all different types of heterogeneity whilst remaining aware of potential confounding. This would however require the use of latent class structures with many different classes and quickly become computationally and empirically infeasible. While we do not suggest that researchers abandon the use of latent class structures for purposes other than taste heterogeneity, we urge for some caution in interpretation and suggest that model averaging can provide a useful tool for checking the likely validity of their insights.

As a closing comment, the findings in the application looking at decision rule heterogeneity are particularly insightful. They suggest that there is more scope for heterogeneity in parameters across individuals conditional on a specific model structure rather than heterogeneity across individuals in the model structure itself. In many ways this is not surprising given that datasets, especially from stated choice survey, are relatively homogeneous in the structure of the choice sets and explanatory variables. The models that work best are more likely to be dataset specific rather than person specific. More work is of course required, including testing using simulated datasets.

This is especially important with a view to looking into the ability of model averaging to uncover heterogeneity of the type analysts increasingly attempt to uncover with latent class structures.

Acknowledgements

References

Abou-Zeid, M. and Ben-Akiva, M. (2010). A model of travel happiness and mode switching. In Hess, S. and Daly, A., editors, *Choice Modelling: The State-of-the-Art and the State-of-Practice*, pages 289–305. Emerald Publishing, UK.

Ben-Akiva, M. and Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1):9–24.

Boeri, M. and Longo, A. (2017). The importance of regret minimization in the choice for renewable energy programmes: Evidence from a discrete choice experiment. *Energy Economics*, 63:253–260.

Börjesson, M., Fosgerau, M., and Algers, S. (2012). Catching the tail: Empirical identification of the distribution of the value of travel time. *Transportation Research Part A: Policy and Practice*, 46(2):378–391.

Bruza, P. D., Wang, Z., and Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. *Trends in cognitive sciences*, 19(7):383–393.

Bureau of Transport Statistics (2012). Household Travel Survey 2010/11: Technical Documentation. *Bureau of Transport Statistics, Transport for New South Wales*.

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

California Department of Transportation (2013). 2010-2012 California Household Travel Survey Final Report.

Campbell, D., Lorimer, V., Aravena, C., and Hutchinson, W. G. (2010). Attribute processing in environmental choice analysis: implications for willingness to pay. 84th Annual Conference, March 29-31, 2010, Edinburgh, Scotland 91718, Agricultural Economics Society.

References

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.

, Claeskens, G. and Hjort, N. L.(2018) Model selection and model averaging. *Cambridge University Press.*

Daly, A. (2010). Cost damping in travel demand models.

Dey, B. K., Anowar, S., Eluru, N., and Hatzopoulou, M. (2018). Accommodating exogenous variable and decision rule heterogeneity in discrete choice models: Application to bicyclist route choice. *PloS one,* 13(11):e0208309.

Fosgerau, M. and Mabit, S. L. (2013). Easy and flexible mixture distributions. *Economics Letters,* 120(2):206–210.

Fox, J. (2015). *Temporal transferability of mode-destination choice models.* PhD thesis, University of Leeds.

Fox, J., Daly, A., and Patruni, B. (2009). Improving the treatment of cost in large scale models. In *European Transport Conference.* Citeseer.

Gazder, U. and Ratrout, N. T. (2015). A new logit-artificial neural network ensemble for mode choice modeling: a case study for border transport. *Journal of Advanced Transportation,* 49(8):855–866.

Gopinath, D. (1995). *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand.* PhD thesis, MIT, Cambridge, MA.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological,* 37(8):681–698.

Hensher, D. A. (2010). Attribute processing, heuristics and preference construction in choice analysis. In Hess, S. and Daly, A. J., editors, *State-of Art and State-of Practice in Choice Modelling: Proceedings from the Inaugural International Choice Modelling Conference,* chapter 3, pages 35–70. Emerald, Bingley, UK.

Hensher, D. A. and Greene, W. H. (2010). Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical Economics,* 39(4):413–426.

References

Hensher, D. A., Rose, J. M., and Greene, W. H. (2012). Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation*, 39(2):235–245.

Hess, S., Daly, A., Dekker, T., Cabral, M. O., and Batley, R. (2017). A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological*, 96:126–149.

Hess, S., Stathopoulos, A., and Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591.

Hess, S. (2014). 14 latent class structures: taste heterogeneity and beyond. In *Handbook of choice modelling*, pages 311–329. Edward Elgar Publishing Cheltenham.

Hess, S., Beck, M., and Crastes dit Sourd, R. (2016). Can a better model specification avoid the need to move away from random utility maximisation? Transportation Research Board (TRB) 96th Annual Meeting.

Hess, S. and Rose, J. M. (2007). *A latent class approach to recognising respondents' information processing strategies in SP studies.* paper presented at the Oslo Workshop on Valuation Methods in Transport Planning, Oslo.

Hess, S., Shires, J., and Jopson, A. (2013a). Accommodating underlying pro-environmental attitudes in a rail travel context: Application of a latent variable latent class specification. *Transportation Research Part D*, 25:42–48.

Hess, S. and Stathopoulos, A. (2012). Linking the decision process to underlying attitudes and perceptions: a latent variable latent class construct. *paper presented at the 13th International Conference on Travel Behaviour Research, Toronto.*

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.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., and Caussade, S. (2013b). It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation*, 40(3):583–607.

Hess, S., Stathopoulos, A., and Daly, A. J. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591.

References

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

Hole, A. R. (2011). A discrete choice model with endogenous attribute attendance. *Economics Letters*, 110(3):203–205.

Ishaq, R., Bekhor, S., and Shiftan, Y. (2013). A flexible model structure approach for discrete choice models. *Transportation*, 40(3):60–624.

Leong, W. and Hensher, D. A. (2014). Relative advantage maximisation as a model of context dependence for binary choice data. *Journal of choice modelling*, 11:30–42.

Montgomery, H. and Svenson, O. (1976). On decision rules and information processing strategies for choices among multiattribute alternatives. *Scandinavian Journal of Psychology*, 17(1):283–291.

Morales, K. H., Ibrahim, J. G., Chen, C.-J., and Ryan, L. M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association*, 101(473):9–17.

Moretti, F., Pizzuti, S., Panzieri, S., and Annunziato, M. (2015). Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing*, 167:3–7.

Outwater, M. L., Bradley, M., Ferdous, N., Bhat, C., Pendyala, R., Hess, S., Daly, A., and LaMondia, J. (2015). Tour-based national model system to forecast long-distance passenger travel in the united states. Technical report.

Posada, D. and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5):793–808.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5):1155–1174.

Scarpa, R., Gilbride, T., Campbell, D., and Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, 36(2):151–174.

Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2010). Probabilistic wind speed forecasting using ensembles and bayesian model averaging. *Journal of the american statistical association*, 105(489):25–35.

References

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.

Swait, J. and Ben-Akiva, M. (1985). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B*, 21(2):91–102.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79:281–299.

Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997). Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(4):433–448.

Walker, J. and Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences*, 43(3):303–343.

Wang, D., Zhang, W., and Bakhai, A. (2004). Comparison of bayesian model averaging and stepwise methods for model selection in logistic regression. *Statistics in medicine*, 23(22):3451–3467.

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., and Kavanagh, R. P. (2003). The use of bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17(6):1579–1590.

Wright, J. H. (2009). Forecasting us inflation by bayesian model averaging. *Journal of Forecasting*, 28(2):131–144.

Chapter 8

Discussion and conclusions

1   Summary

This thesis has made a number of theoretical contributions as well as providing a number of detailed empirical applications in the context of the interface between mathematical psychology and mainstream choice modelling. In the introduction, a number of research gaps were identified to bridge the gap between these disciplines. This chapter starts by mapping the relevant contributions of each chapter with the research gaps detailing the advances that have been made. It then reviews the conclusions that can be drawn by considering the chapters together.

Gap 1: Tractable methods for the applications of accumulator models.

The travel behaviour community has seen a considerable rise in the adoption of models incorporating further behavioural insights. However, most efforts thus far (such as random regret minimisation, Chorus 2010) are mathematically not very different in terms of their underlying structure compared to traditional random utility maximisation models. This thesis considers the argument that if we are to truly incorporate behavioural insights, then we should move further afield. In particular, 'accumulator' or 'process' models developed within mathematical psychology are often also designed to explain and predict multi-attribute, multialternative choices (Roe et al., 2001; Trueblood et al., 2014) making them strong candidates for the transition into mainstream choice modelling. In the introduction, we identified a number of key issues that have thus far limited the application of these models, the most severe of which being the computational complexities of the models (Otter et al., 2008) as well as the lack of a clear basis for best practices with which to implement them.

In Chapter 2, we address a key methodological issue for decision field theory (DFT). Previous applications of DFT have either been largely theoretical or have made limiting assumptions, such as setting the number of preference updating steps within the model to a very large number (Berkowitsch et al., 2014; Trueblood et al., 2014). We however demonstrate that the probability with which alternatives are chosen can be simply calculated for any number of preference updating steps. As well as leading to greater model flexibility, this allows for many crucial additions in the context of travel behaviour research, such as the incorporation of initial preferences or underlying biases. Furthermore, Chapter 2 illustrates a number of possible scaling methods to deal with DFT's scale invariant nature, although it is in Chapter 4 that we find the best solution for this through the adoption of scaling parameters whilst fixing the attention weight parameters. The main benefit of this is that a priori knowledge for the directionality of the different attributes is no longer required. Additionally, we explore the nature and impact of the different 'process' parameters in DFT, testing possible normalisations that can be performed to improve the robustness of DFT. Chapter 3 in particular takes an in-depth look at the precise nature of the parameter for the number of preference updating steps. We demonstrate that it performs a role similar to a scale parameter within a random utility model. An additional key benefit of the work that has been conducted on DFT in this thesis is that it has led to the development of free to use, easily adaptable code for modelling choices under a DFT model, which has been implemented as part of the Apollo choice modelling package (Hess and Palma, 2019).

We also provide a detailed operationalisation of the multi-attribute linear ballistic accumulator (MLBA) model (Trueblood et al., 2014) in Chapter 4, a model that has never been tested in the context of travel behaviour analysis. Together with DFT, we demonstrate that both models can incorporate alternative and attribute specific coefficients, income effects and socioeconomic differences. As with DFT, we provide a detailed discussion on the different process parameters within MLBA. Whereas previous applications of MLBA have often just fixed parameters arbitrarily, we provide mathematical insights for normalisation as well as thorough empirical tests for the identification of parameters. For both DFT and MLBA, we find that in addition to theoretical identification, the empirical identification of the process parameters is dataset specific, with additional normalisations required for simpler datasets.

Finally, we also provide first steps towards introducing decision field theory to dynamic data settings in Chapter 6. Given that accumulator models are dynamic in nature, it is somewhat surprising that this idea has been rarely investigated before, but we demonstrate a number of important first steps for the application of dynamic models in general in dynamic settings. As well as further operationalising DFT such that it can incorporate changing attribute values, we also show how to adapt DFT in the context of driver behaviour

modelling.

Gap 2: Rigorous comparisons of alternative approaches

Whilst (Berkowitsch et al., 2014) provides a thorough comparison of (a restricted version of) DFT against multinomial logit and probit in the context of consumer choices, DFT as well as MLBA have rarely been tested or at all on stated preference or revealed preference studies within travel behaviour modelling. In Chapter 2, we demonstrate that DFT performs well in the context of route choice, can incorporate underlying preferences towards alternatives, and can provide 'relative importance of attributes' measures. Whilst these are not equivalent to standard econometric welfare measures, we demonstrate that the outputs from DFT in line with those of traditional models and give an insight into preferences. Crucially, in applications in both Chapter 2 and Chapter 4, we find that DFT outperforms traditional models in terms of model fit. There is little difference in the performance between DFT and MLBA, although our further extensions to DFT in Chapters 3 and 6 suggest that it may have more scope for further operationalisation within travel behaviour modelling. Chapter 4 additionally applies both models to rigorous tests in simulated settings. We find that both models provide stable parameter recovery as well as providing enough flexibility to additionally perform well for datasets where the choices were generated by either a standard multinomial logit model or a random regret minimisation model. These tests also allowed for detailed explorations into the impacts of fixing process parameters within DFT and MLBA, with it being clear that DFT's feedback parameters do not always have an impact and can often be fixed without impacting model performance.

Further key empirical applications in this thesis have looked at bringing both DFT and MLBA outside of experimental conditions. With both models typically utilised on laboratory based choices, the lessons learnt from the applications of DFT and MLBA for stated preference data and simulated data in this thesis allowed us to apply them both to a revealed preference dataset for (as far as we are aware) the first time. Our application demonstrated that both DFT and MLBA provided good performance in both estimation and in forecasting with subsets of validation samples. Furthermore, in Chapter 6 we demonstrated that DFT could be further adapted in the context of dynamic driving behaviour. Whilst this work is preliminary in nature, it demonstrated that DFT is by no means only for experimental data, with future applications of it likely to have a large impact if further developments are explored. Over the course of the thesis we test DFT across route, accommodation, conservation, simulated, RP and dynamic choice contexts, demonstrating that good performance is not specific to a particular choice scenario.

Gap 3: Developing a choice modelling framework based on quantum logic

Another gap identified for further exploration is to test whether models with a quantum framework can make a successful transition into mainstream choice modelling as they have done in cognitive psychology. As a key theme of work in quantum cognition has been based on the explanation of ordering and interference effects, both of which are regularly observed and considered within choice modelling contexts, it is unsurprising that first steps have been made through Lipovetsky (2018)'s quantum work considering consumer choices and Yu and Jayakrishnan (2018)'s work demonstrating that quantum logic can provide a useful transition from preferences under stated preference settings to preferences under revealed preference settings. In Chapter 5, we discuss in detail how quantum logic works as well as providing demonstrations of how it can be applied in a choice modelling context. We also discuss a number of important requirements for a quantum model framework before introducing two possible approaches. The first of these incorporates regret-like functions whilst the second utilises concepts from MLBA models. Chapter 5 then empirically test both versions as well as a quantum model based on trigonometric functions as specified by Lipovetsky (2018). We find that quantum models provide good model fit both in estimation and in forecasting with subsets of validation samples.

Furthermore, work in Chapter 5 also explores the concept of 'quantum rotations', a simple mathematical construct that allows for a 'change in perspective' through an adjustment of the projection lengths which define the probabilities with which alternatives are chosen. We find that these rotations greatly improve the flexibility of quantum models and seem to provide a good account of behaviour. Additionally, Chapter 6 also explores an application of quantum models in the context of driving behaviour, demonstrating how the model can be adjusted appropriately for such behaviour. As with DFT, the thesis as a whole illustrates that quantum models can be applied across multiple scenarios with varying characteristics.

Gap 4: Identification of contexts for which accumulator and quantum models are suitable

Crucially, given that neither decision field theory or the multi-attribute linear ballistic accumulator model can provide welfare measures, their application in mainstream choice modelling approaches is limited to contexts in which these measures are not required. This is of course the same for other non-RUM models but have not stopped their use for other purposes.

In the context of predicting behaviour, we show numerous examples of both models outperforming typical choice models in both Chapters 2 and 4 in terms of model fit. Whilst further extensions for incorporating heterogeneity are clearly required, Chapter 2 importantly demonstrates that DFT can incorporate random parameters.

In Chapter 6, we also identify dynamic choice settings as a crucial further setting for testing accumulator models such as DFT. Given that these models are already designed to capture the process of decision-making, they intuitively should provide a natural framework in the context of dynamic data. Initial applications of DFT in the context of driving behaviour illustrate that there is a lot of scope for further developments in this regard, with our DFT model giving similar outputs to traditional models in, for example, how drivers treat time headways and with DFT performing comparably in terms of model fit.

In Chapter 5 we also test the quantum models that we developed across a number of choice contexts, including testing quantum models in the context of moral decision-making. However, it is not clear that their performance here is particularly different to their performance in our route choice applications. It thus remains to be seen where these models are likely to have their greatest impact.

Gap 5: Testing model averaging and combining approaches from different disciplines

As well as contrasting and comparing models developed in mathematical psychology in comparison to standard choice models, the introduction of this thesis also identifies a gap in terms of testing the merit of the use of model averaging within choice modelling as well as a lack of applications using approaches from different disciplines together. In Chapter 7, we demonstrate across a number of contexts the potential benefits of model averaging approaches. Firstly, we illustrate that model averaging can be implemented effectively for contexts where the choice of a final model is difficult. Secondly, it also performs well in the context of large-scale datasets for which there is too much computational burden for state-of-the-art choice models, but simpler models are unsatisfactory. Consequently, with our results demonstrating the effectiveness of model averaging across a large set of candidate models, it is clear that we can also average across accumulator and quantum models as well as traditional models, with further work in Chapter 7 finding an improved model fit found by averaging across latent class models with different decision rules. Additionally, the use of both types of models together can help inform analysts what the key sources of heterogeneity in their dataset may be.

2   Objectives and contributions

In the introduction of this thesis, we identified six distinct objectives for bridging the gap between mathematical psychology and econometrics. The following section summarises how the work in this thesis has approached and addressed these methodological (M1-M3) and applied (A1-A3) objectives.

M1: The operationalisation of decision field theory (DFT) for choice modelling applications: This objective was met by developments made in Chapter 2, where we improved the underlying mathematics allowing for an easier calculation of the likelihood function for DFT. Additionally, this chapter as well as Chapter 4 provided further steps towards improving the flexibility of DFT, through an improved method for scaling attributes and illustrations of how DFT can incorporate a number of typical factors considered in standard choice modelling, such as the incorporation of mixed parameters, alternative specific parameters, attribute specific parameters, income and sociodemographic effects. Chapter 3 additionally developed methods for the incorporation of the effects of choice task response time within DFT and Chapter 6 further extended DFT for application in driving behaviour analysis.

M2: The operationalisation of the multi-attribute linear ballistic accumulator (MLBA) model: This objective was met through developments made in Chapter 4. As with DFT, we demonstrate that MLBA can be adjusted to include attribute and alternative specific constants as well as sociodemographic effects. This chapter also provides a number of steps considering the normalisation of process parameters within MLBA such that standard practices for the application of MLBA can be established.

M3: The development and operationalisation of choice models based on quantum logic: This objective was met through work conducted in Chapter 5, in which we develop two distinct new frameworks for the incorporation of quantum logic. The first is based on ideas from random regret minimisation and the second based on ideas from the multi-attribute linear ballistic accumulator. We additionally develop the concept of quantum rotations for capturing changes in choice context and test this approach across a number of choice contexts in Chapter 5. We demonstrate the effectiveness of quantum rotations applications at capturing simple contextual changes such as the effect of the order of attributes and alternatives, as well as potentially more complex changes in perspective in the moral context of taboo trade-offs and choices that impact yourself and your partner.

A1: To rigorously test models from mathematical psychology against mainstream choice models: This objective was met through applications across a number of chapters within this thesis. Firstly, DFT was rigorously compared to both multinomial logit and random regret minimisation models in Chapter 2, before DFT as well as MLBA were compared against these models in Chapter 4. These models were tested on a large variety of datasets, including stated preference, revealed preference and simulated datasets. Whilst the performance of the models is dataset-specific, we find that in general these models perform comparably to standard choice models. Furthermore, DFT was tested on housing, accommodation and route choice in Chapter 3 and on driving behaviour in Chapter 6.

A2: To test accumulator and quantum models in the context of real world choices: This objective was met through the work conducted in Chapter 4 as well as the work in Chapter 6. In Chapter 4, we test DFT and MLBA in the context of consumer preferences regarding different providers for train journeys to London. Both DFT and MLBA performed well in terms of model fit, as well as providing interesting insights in terms of the relative importance of different attributes. In Chapter 6, we provide (as far as we aware) the first test of both DFT and quantum models in the context of lane merging choices.

A3: To explore the insights generated by applications combining these models: This objective was met through the work conducted in Chapter 7, which first establishes the benefits of applying model averaging across a large number of choice models. We then demonstrate the use of applying latent class models with a wide range of pairs of models in the different classes, showing how model averaging can be used to highlight possible confounding sources of heterogeneity.

3 Outlook

This thesis has discussed a number of important methodological developments for the use of accumulator and quantum choice models in the context of travel behaviour research, but we have also highlighted a number of steps for future research at the end of each chapter. In this section, we summarise these steps with respect to the research gaps that were discussed in the previous sections.

Firstly, whilst the work in this thesis provides a number of theoretical developments and offer some proof-of-concepts, there is much work left still to do if accumulator or quantum models are to be fully established in mainstream choice modelling. Whilst we demonstrate that random parameters

can be incorporated for DFT, further tests could implement random parameters for MLBA and quantum. Additionally, Terry et al. (2015) demonstrate that different assumptions about the types of distributions used within the context of the linear ballistic accumulator model can be made, meaning that the introduction of this increased flexibility is likely also possible at the multi-attribute level. More crucially, given that these models do not provide welfare measures and have been predominantly tested in terms of model fit in this thesis, one could argue that we should instead be adopting machine learning approaches rather than adding further theories and complexities to our models. An initial challenge that could be addressed in this regard is the comparison of results from machine learning against model averaging across a wide range of candidate models including models developed and refined in this paper. Furthermore, additional testing of model outputs is required in terms of understanding, for example, whether elasticities and forecasts that are generated by accumulator models make sense. Alternatively, interesting future steps could consider more specific tests regarding the prediction of future behaviour or backcasting to test the reliability of the transferability of accumulator models in comparison to traditional choice models and machine learning. The behavioural foundations of the models from mathematical psychology should provide clear benefits here.

Secondly, given that accumulator models are designed to capture the underlying choice process as well as predicting the choice outcomes, they should be tested as process models by moving beyond the context of choice only datasets. This could provide a more detailed examination of the process parameters within both DFT and MLBA. Whilst the work in Chapters 3 and 6 makes a start on this topic, further insights could be generated through testing whether DFT and MLBA can predict response times as well as choice outcomes in the context of travel behaviour, as often done within mathematical psychology applications. Furthermore, they should additionally be tested with, for example, eye-tracking or neuroimaging data, which may help utilise the full flexibility of the models (rather than simply fixing various parameters). Such tests, as well as applications of the models in the presence of contextual effects (which may have different results with the refinements made in this thesis) will help establish the precise mathematical impacts of these parameters and may possibly differentiate DFT and MLBA, with our applications in Chapter 4 showing little difference in model performance. Future applications of quantum models could also consider more detailed data in the context of, for example, moral machine type dilemmas, which could help generate further insights and allow us to draw more definite conclusions with regards to the type of context quantum models are most suited to. Additionally, whilst we briefly consider large-scale datasets in the context of model averaging across basic models, none of the models developed and refined in this thesis are tested under such scenarios, although future

work would also need to reduce the computational burden of these models to achieve this step.

Another step that is required is to make more progress in ensuring these models are more applicable and hence further establishing standard practices for the implementations of these models. Whilst we provide a number of important steps in this regard in Chapter 4, the best normalisations of DFT, MLBA and quantum models are dataset specific, and thus it takes more work to establish the best version of the model than it does for standard choice models. Adjustment of the precise structures of these models may make them more approachable if, for example, it was possible to run a generic version of the model without risk of confounding parameters or overspecification.

Finally, there is a vast amount of scope for further developments for accumulator models in dynamic choice settings. With a number of possibilities for the increased flexibility of dynamic models in dynamic settings discussed in Chapter 6, it is clear that the empirical application is this chapter is just a first step. Different settings may reveal very different results, with there not being much differentiation between alternative models in the context of lane merging choices, likely as a result of little variation in behaviour observed. Work is also needed on exploring appropriate approaches for determining how often updated external information is internalised by a decision maker.

Overall, the work from this thesis illustrates that models and ideas from mathematical psychology applied in the context of mainstream choice modelling are flexible, adaptable and provide an interesting alternative to standard choice models. It is with a significant level of confidence that I conclude that further attempts to bridge the gap will undoubtedly bring a richer understanding of choice behaviour, for mathematical psychologists and mainstream choice modellers alike.

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.

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

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

Lipovetsky, S. (2018). Quantum paradigm of probability amplitude and complex utility in entangled discrete choice modeling. *Journal of choice modelling*, 27:62–73.

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.

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.

Terry, A., Marley, A., Barnwal, A., Wagenmakers, E.-J., Heathcote, A., and Brown, S. D. (2015). Generalising the drift rate distribution for linear ballistic accumulators. *Journal of Mathematical Psychology*, 68:49–58.

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.

Yu, J. G. and Jayakrishnan, R. (2018). A quantum cognition model for bridging stated and revealed preference. *Transportation Research Part B: Methodological*, 118:263–280.

Appendix A

Appendix to Chapter 2

1 The impact on including socio-economic factors on the differences between DFT and MNL

Figure A.1 shows the impact of including socio-economic factors (income effects) on the difference between MNL and DFT predicted probabilities of the chosen alternatives for SP-1. Grey points indicate decisions made by individuals who earn less than 80,000 Swiss Francs, whereas black points indicate decisions made by individuals who earn more. It appears that there are no significant differences between the models by including these income effects, despite the fact that the MNL model improved in model fit more significantly.
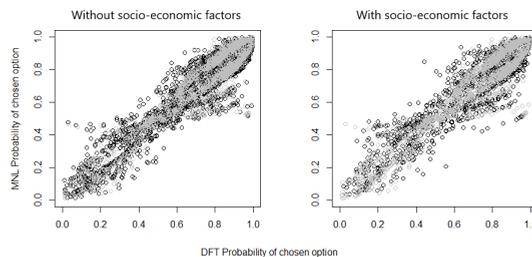
Fig. A.1: Impact of including income on MNL and DFT models on dataset SP-1

2 DFT model estimates

The weights used for SP-2 are travel time (TT), cost (TC), rate of crowded trips (CT), rate of delays (RD) and the average length of delays, where this final weight is fixed such that the weights together sum to one. The cost of

Swiss (SP-1)

	Model 1		2		3		4		5	
Model	1		2		3		4		5	
time par.	yes		no		yes		yes		no	
weights	fixed		fixed		truncated		truncated		truncated	
other pars.	fixed		fixed		normal		truncated		truncated	
pars.	7		6		14		14		12	
LL	-1,597.30		-1,608.65		-1,450.39		-1,438.39		-1,430.41	
BIC	3,252		3,266		3,015		2,991		2,959	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
$\mu_{wt_{TT}}$	0.3480	45.99	0.3353	34.40	0.3406	8.69	0.3382	13.24	0.3360	12.42
$\mu_{wt_{TC}}$	0.4691	43.20	0.4581	54.45	0.5264	7.76	0.5202	11.14	0.5286	10.41
$\mu_{wt_{HW}}$	0.0739	13.03	0.0841	51.89	0.0598	12.29	0.0612	12.74	0.0597	13.02
μ_{ϕ_1}	-0.01	-0.02	1.73	109.57	3.5451	4.23	2.0106	7.46	184.649	4.94
μ_{ϕ_2}	0.00	0.01	0.16	48.97	-0.0986	-1.93	-0.0301	-0.74	0.0916	8.01
μ_e	0.00	0.00	0.28	5.68	0.0001	0.19	0.0027	3.25	0.0015	6.32
μ_t	10.05	12.14	-		29.8485	7.40	29.3096	9.49	-	
$\sigma_{wt_{TT}}$	-		-		0.1875	3.57	0.2019	5.16	0.1472	3.30
$\sigma_{wt_{TC}}$	-		-		0.3894	3.45	0.4096	5.35	0.3836	4.23
$\sigma_{wt_{HW}}$	-		-		0.0393	6.42	0.0405	6.83	0.0356	7.40
σ_{ϕ_1}	-		-		1.0818	3.28	0.8760	3.95	41.2838	3.98
σ_{ϕ_2}	-		-		0.1720	2.57	0.0757	4.28	0.0954	4.65
σ_e	-		-		0.0019	2.68	0.0013	2.81	0.0009	5.77
σ_t	-		-		9.3739	5.19	5.7655	3.20	-	

Table A.1: Results for SP-1

	UK (SP-2)									
Model	1		2		3		4		5	
time par.	yes		no		yes		yes		no	
weights	fixed		fixed		truncated		truncated		truncated	
other pars.	fixed		fixed		normal		truncated		truncated	
pars.	8		7		16		16		14	
LL	-3,598.87		-3,676.34		-3,156.27		-3,140.09		-3,190.23	
BIC	7,263		7,410		6,444		6,412		6,495	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
$\mu_{wt_{TT}}$	0.1132	5.33	0.1350	11.63	0.1001	7.29	0.1082	7.46	0.0854	10.29
$\mu_{wt_{TC}}$	0.6786	12.72	0.6642	41.98	0.7889	9.53	0.7802	9.41	0.8205	8.85
$\mu_{wt_{CT}}$	0.0691	3.30	0.0816	10.72	0.0339	4.09	0.0320	2.78	0.3361	6.38
$\mu_{wt_{RD}}$	0.0271	3.97	0.0308	3.39	0.0142	2.88	0.0191	5.27	0.0122	2.09
μ_{ϕ_1}	0.1000	13.68	152.5165	102.80	0.0840	3.56	0.0742	5.19	0.1854	3.86
μ_{ϕ_2}	1.9381	180.09	0.4283	9.99	0.7900	12.09	0.8091	7.40	0.6418	4.61
μ_ϵ	0.0393	8.48	0.1543	5.31	0.0300	10.63	0.0448	5.06	0.0872	3.34
μ_t	3.7792	242.14	-		10.7987	14.41	10.1687	9.25	-	
$\sigma_{wt_{TT}}$					0.0623	4.21	0.0746	5.35	0.0575	5.81
$\sigma_{wt_{TC}}$					0.8821	9.63	0.7816	8.63	0.8988	6.50
$\sigma_{wt_{CT}}$					0.0616	3.12	0.0738	5.50	0.0311	6.54
$\sigma_{wt_{RD}}$					0.2486	4.35	0.0298	4.33	0.0179	5.31
σ_{ϕ_1}					0.0717	5.09	0.0273	4.80	0.0907	3.81
σ_{ϕ_2}					0.4196	10.26	0.3342	7.98	0.4683	4.91
σ_ϵ					0.0131	9.09	0.0527	4.12	0.0822	2.75
σ_t					3.7301	7.16	3.2715	5.85	-	

Table A.2: Results for SP-2

a provision of a delay information service was found to be insignificant and therefore omitted.

3   Notes on DFT parameters

It should be noted that large differences in ϕ_1 may not have much impact on a decision. For example, suppose we have the following choice task:

	Attribute 1	Attribute 2	Attribute 3
Alternative A	3	4	5
Alternative B	2	4	6
Alternative C	3	7	1

Table A.3: An example choice task

If at some time point we had a preference vector of $P_t = [10, 9, 8]'$ and a value of 0.05 for ϕ_2, then the following results would be obtained for $S \times P_t$ for the given values of ϕ_1:

ϕ_1	$P_t[1]$	$P_t[2]$	$P_t[3]$
0.1	9.11	7.99	7.09
0.5	9.48	8.45	7.50
1	9.50	8.53	7.58
10	9.50	8.55	7.60

Table A.4: Impact of feedback parameters on probabilities

This means that the difference in preference between alternatives is not much impacted by ϕ_1. This could particularly be the case for choice scenarios involving only two alternatives, as shown by the minimal impact adjustments on ϕ_1 and also ϕ_2 had on SP-1 (see Table 2.9). Future work on DFT could look at the impact of removing these parameters altogether.

Additionally, we can also use this choice task to demonstrate how the timestep and error parameters, t and ϵ capture distinctly different features of the data. The table below gives the probability of choosing the three alternatives when $wt_1 = 0.3$, $wt_2 = 0.3$, $wt_3 = 0.4$, $\phi_1 = 0.1$ and $\phi_2 = 0.05$:

Under these conditions, the expected valence, $\mu = [0.3, 0.45, -0.75]'$. This means that with more timesteps, we would expect stronger preferences towards alternatives A and B. Higher values for the number of timesteps indicates that the decision-maker is more likely to consider all of the attributes. This results in the variance of the attribute weights having less impact. Higher values for the error variance ϵ result in the relative differences between

Probability of choosing alternatives				
t	10	20	10	20
ϵ	1	1	5	5
Alternative A	0.2807	0.2933	0.3449	0.3628
Alternative B	0.5265	0.5811	0.4721	0.5158
Alternative C	0.1928	0.1255	0.1830	0.1214

Table A.5: Impact of process parameters on probabilities

attributes being less significant. (For example, $\epsilon = \infty$ results in all alternatives being chosen with equal probability). Another way of considering these two parameters psychologically is that they are 'quality' and 'quantity' of information processed. The number of timesteps tells you how much of the information is considered (hence lower values imply less predictable choices) and the error variance tells you how 'distinct' the decision-maker interprets the alternatives (with high values meaning that the decision-maker interprets there being little difference between the alternatives).

Finally, this example can also be used to demonstrate that the scale-variant nature of DFT arises from the variance of the weights. If, for example, attribute 3 values were doubled, then to obtain an equivalent expected valence of $\mu = [0.3, 0.45, -0.75]'$, the weight for attribute 3 would need to be decreased relative to the weights for attributes 1 and 2. Weights of $wt_1 = 0.375$, $wt_2 = 0.375$ and $wt_3 = 0.25$ achieve an expected valence of $\mu = [0.375, 0.5625, -0.9375]'$, exactly 1.25 times the previous μ. However, this would result in a very different value for Ψ, as the variance of the weights has changed. Consequently the probabilities of alternatives would change, despite the relative expected valences remaining the same.

4   Models results from Hess et al. (2016)

UK data:	MNL	RRM	mu-RRM	mixed MNL	mixed RRM	mixed mu-RRM
LL	-3,721.67	-3,699.49	-3,698.89	-3,184.89	-3,205.27	-3,174.96
parameters	7	7	8	12	12	13
BIC	7,500.81	7,456.46	7,463.47	6,468.30	6,509.08	6,456.66
Runtime (normalised)	1.00	1.58	2.11	50.75	316.98	335.69

Table A.6: Mixed model results for SP-2 from Hess et al. (2016)

Appendix B

Appendix to Chapter 4

1 $MLBA_0$

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., 2013) 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:

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

where v_j is a valence function and γ is a logistic parameter. Small values of the logistic parameter γ 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

$$V = C \cdot M \cdot W \tag{B.2}$$

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:

$$C_{i,j \neq i} = \frac{exp(-\phi \cdot Dist_{i,j}) - 1}{n - 1}. \tag{B.3}$$

Finally, ϕ 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 B.1 below.

Table B.1: 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.