Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity

Andrew T. Collins

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Transport and Logistics Studies
University of Sydney Business School

December, 2012

Abstract

An extensive literature has recognised that when discrete choices are made, only a subset of the attributes of the choice alternatives may be considered or attended to by each decision maker. The wider literature suggests that attribute nonattendance (ANA) is important, and that failure to recognise ANA contributes to biased model outputs such as willingness to pay measures, masked sensitivities, implausibly signed random parameters coefficients, and exaggerated taste heterogeneity. It may also be of intrinsic interest to the analyst, and reveal problems with stated choice experimental designs. This research uses simulated data to gain a deeper understanding of the biasing influences of ANA. It is shown that random parameters logit models handle ANA poorly, with the extent of the bias in the model outputs driven by both taste heterogeneity and ANA. The literature has identified some shortcomings and limitations of the existing methodologies for handling ANA. The simulated data are employed to further critique these methodologies, and demonstrate that they are likely to introduce their own biases.

This thesis proposes the random parameters attribute nonattendance model, and seeks to improve upon the existing methodologies. The model combines discrete and continuous random parameters, and can infer ANA and taste heterogeneity, without the need to collect supplementary data. The model is tested on simulated data with encouraging results. In addition, in an empirical setting of short haul flight choice, with real stated choice data, the model outperforms the RPL model and several existing ANA methodologies. Necessary conditions to ensure identification are discussed. The model allows for a balance between parsimony, and the handling of correlation of ANA, through a spectrum of possible model specifications; with this tension explored in detail. Further insights into ANA behaviour are gained in the empirical study. The thesis makes a useful methodological contribution, by developing a model with unique properties that can capture the important and prevalent behaviour of ANA.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Andrew T. Collins

Acknowledgements

My first acknowledgements are to my supervisors, Professors John Rose and David Hensher. I have known them both for nearly ten years now, and in that time they have assumed many roles: supervisors, yes, but also bosses, colleagues, coauthors, business partners, and mentors. In each of these roles I have gained much, and for that I thank them. John has been instrumental to the progression of my candidature. He has encouraged me to think of all the possibilities, and provided me with the knowledge and mindset to turn those possibilities into realities, from the ground up. His ideas are always flowing, and his feedback insightful. I particularly appreciate David's relentless energy and enthusiasm, and of course his immense experience and ability. I would also like to thank Professor Michiel Bliemer, who provided valuable feedback on my notation.

It is my colleagues at ITLS that make it such a fantastic place to work. I'd especially like to thank the coffee group for all the fun and banter - John, Matt, Geoffrey and Michiel. Special mention to the coffee machine too, which really can be seen as an embodiment of David's benevolence. To my friends beyond ITLS, I'm back! I'm immensely looking forward to Wednesday night climbing, weekends away, and the travel we have planned. In particular, I'd like to thank Kenny and Keo for their great friendship over the years, and Fruzsina, my awesome friend and housemate.

I have no doubt that some of the strongest memories of my candidature will be the many Saturday mornings spent with my sister Catherine, walking the coastline, discussing academia, and searching for the best bacon and egg roll! My brother Anthony has been coconspirator for some long, epic day walks, and a great host in New York, along with Angela and Emelia. My parents have provided unerring support and encouragement, and have welcomed me every time I have visited the farm. Certainly the farm has been a fantastic location for working towards my PhD, through all the stages. I thank my father Roger, for his big picture insights on work, academia, and life, and the wonderful distractions talking about and drinking wine. My mother Mary has provided endless hospitality for which I am truly grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Since their inception, econometric models of discrete choice, such as the conditional logit model (e.g., McFadden, 1974), have provided the analyst with a powerful mechanism for analysing the choice of discrete outcomes. Part of the appeal of such models stems from their Lancasterian nature (Lancaster, 1966), whereby a choice alternative can be viewed as a bundle of utility bearing attributes. When this approach is applied within a random utility framework (Marschak, 1960), the probability of an alternative being chosen can be expressed as a function of these attributes. Also noteworthy are the range of outputs that such models can provide, including choice share predictions, direct and cross elasticities, and marginal rates of substitution, most commonly represented as willingness to pay (WTP) for changes in specific attributes of choice alternatives. Such versatility has allowed choice models to be applied in such diverse fields as transportation, marketing, health economics, environmental economics, recreation and tourism demand and energy economics. The model outputs frequently guide both policy and market decisions. Consequently, the econometric and behavioural soundness of such models is of key importance. In particular, to maintain confidence in the results, the assumptions of the models need to be sufficiently plausible in the context in which they are applied.

Fortunately, discrete choice models have proven to be rather flexible, and a number of assumptions of the conditional logit model (frequently referred to as the multinomial logit (MNL) model) that lead to behaviorally implausible outcomes have been relaxed over the years. An important advancement has been the nested logit model (Ben-Akiva, 1973; Daly and Zachary, 1978; Williams, 1977; McFadden, 1978), and, more broadly, the generalised extreme value family of models, which relax the MNL model's assumption of independence from irrelevant alternatives. A further assumption of the basic MNL model is homogeneity of preference across a sample for any variable entered into the model. While heterogeneity in preference for the attributes of the choice alternatives can be captured systematically through their interaction with socio-demographic information and other covariates, these covariates may not be available to the analyst, and in any case it is behaviorally plausible for preferences to vary across individuals stochastically. Individual level models are usually not feasible due to a limited quantity of data at the individual level. However, random parameters logit (RPL) models (Train, 2009; Hensher and Greene, 2003, provide excellent overviews) allow such stochastic variation in preferences to be captured, and consequently have seen widespread acceptance. Latent class (LC) models (Kamakura and Russell, 1989) also overcome the assumption of preference homogeneity, by identifying numerous preference segments, where preferences may vary across segments.

The above extensions to the MNL model have focused on the assumed structure of the unexplained components of utility. Another research stream has focused on the process by which decision makers come to make a choice. This has been motivated by extensive evidence in the psychology literature that information can be processed in many different ways, using a variety of heuristics, prior to and during the making of a choice (see for example Tversky, 1972; Payne et al., 1992). By contrast, most econometric models of choice assume that the attributes of the choice alternatives are processed in a fully compensatory fashion. That is, all attributes of the choice alternatives are attended to. Some studies have considered, within the discrete choice modelling framework, deviations from the fully compensatory choice process. Examples include choice on only a single attribute (lexicographic choice; Saelensminde, 2006), aggregation of common-metric attributes (Layton and Hensher, 2010), and the avoidance of negative emotions (regret-minimisation; Chorus et al., 2008). Research into choice heuristics and decision rules is extensive and ongoing (for a review, see Leong and Hensher, 2012).

One choice heuristic that has received particular attention in the literature in recent years is the ignoring of attributes, commonly referred to as attribute nonattendance (ANA). This stream of research recognises that any individual, on any choice occasion, might only attend to a subset of the available attributes that describe a choice alternative. For example, Hensher et al. (2005) and Rose et al. (2005) asked respondents in stated choice (SC) experiments whether they ignored specific attributes. They found non-trivial rates of ANA, and that accounting for nonattendance led to significantly different WTP measures. Attribute nonattendance has been studied in many fields, including transportation (e.g., Hensher et al., 2005), environmental economics (e.g., Campbell et al., 2011), marketing (e.g., Gilbride et al., 2006), and health economics (e.g., Hole, 2011a) contexts. Several prevalent themes and unresolved questions have been established. There is a strong focus on the implications of ANA, specifically any bias in WTP that results from failing to account for ANA (Hensher et al., 2005), and the implications for measures of WTP of respondents not attending to a price attribute (Scarpa et al., 2009). Studies that utilise the respondents' stated nonattendance typically note the potential unreliability of such responses (e.g., Hensher et al., 2007). An alternative approach, that is not reliant on supplementary data, is to infer ANA analytically, through an appropriate specification of the choice model (Swait and Adamowicz, 2001; Train and Sonnier, 2005; Gilbride et al., 2006; Hess and Rose, 2007; Hess and Hensher, 2010).

Motivated by the problems that may stem from not handling ANA, an incomplete knowledge of the extent of these problems, and the limitations of existing techniques for detecting and accommodating ANA, this thesis investigates the impact of ANA in detail, and introduces a new method for inferring ANA. The next section examines these motivations in more depth.

1.2 Motivation for the thesis

This section will first broadly consider why ANA is important, and why research in the area, including that of this thesis, is warranted at all. Then, some specific gaps in the literature will be detailed. The next section will summarise the contribution of this thesis to the literature.

As discussed in the previous section, choice models typically rely on the assumption of fully compensatory behaviour across all attributes of the choice alternatives. Failure to account for individuals who do not attend to subsets of the attributes has been shown to bias key model outputs, including WTP (Hensher et al., 2005). Indeed, there is an extensive literature that examines possible sources of bias in WTP measures, including hypothetical bias in SC data (Hensher, 2010), choice task complexity (Deshazo and Fermo, 2002), attribute ordering within the choice task of the SC experiment (Kjaer et al., 2006) and lexicographic choice (Saelensminde, 2006). The extensive efforts to identify and mitigate biases in the WTP are consistent with the severity of the problem. Suboptimal policy, pricing, and product development decisions may result. Non-market valuation applications, common in fields such as environmental economics, are particularly dependant on WTP measures. Bias is not the only concern. The question arises of how to treat nonattendance to a price attribute, as this precludes identification of a WTP for nonattending individuals. Worryingly, Scarpa et al.

(2009) found that 80 to 90 percent of respondents did not attend to the price attribute in a study designed to provide valuations for rural landscape. More positively, accounting for ANA in this study led to lower and more realistic WTP estimates. Consequently, having robust methods to identify ANA may mitigate its biasing influence.

Attribute nonattendance may be a consequence of poor experimental design, for example, where prices are too low or too high. At the very least, a robust mechanism for handling ANA would curb the biasing influence of ANA induced by the design. By identifying, with this mechanism, problems stemming from the experimental design, an effort could be made to try and improve the design and reduce the incidence of ANA.

One interpretation of ANA is that it is not an experimental artefact, but rather a valid phenomenon reflecting the preferences of the individual making a choice. Cirillo and Axhausen (2006) suggested that some automobile drivers might legitimately have a zero, rather than negative, valuation of time in the vehicle, where an inflated mass at zero exists alongside some distribution of negative valuations. Gilbride et al. (2006) recognised that consumers choosing a product might have no intrinsic value for some of the attributes of the products on offer. In contexts such as these, ANA is no longer just a problem that may bias overall estimates, but an expression of preference heterogeneity, and a valid behavioural phenomenon that may be of interest to the analyst. Again, however, robust methods for identifying ANA are important.

Numerous other motivations for handling ANA have been identified in the literature. Sensitivities which may otherwise be insignificant may be revealed as significant for the subsample that attends to an attribute (Rose et al., 2005). The incidence of implausibly signed coefficients in RPL models may be reduced (Hensher, 2007). Finally, failing to handle ANA may result in overinflated measures of preference heterogeneity (Hensher, 2007; Puckett and Hensher, 2008; Campbell and Lorimer, 2009; Hess and Hensher, 2010).

Despite an extensive and flourishing literature, gaps remain. There is an acknowledgement in the literature that ANA that has not been accounted for is likely to bias the magnitude of the associated parameter downwards. Similarly, there is some evidence to suggest that ANA biases the measures of dispersion of random parameters upwards, and that this may be one cause of the often observed phenomenon of implausibly signed coefficients (Hensher, 2007). However, most examples of these three findings in the literature are constrained by an empirical setting, which only allows a limited number of data points to be observed, and which raises difficulties in measuring bias, as the truth cannot be known with certainty. There remains a lack of precise knowledge about the nature of these biases and sign violations, in terms of their severity, the direct influence of the ANA rate, and the interaction of ANA with other phenomena such as taste heterogeneity.

While the literature has not resolved whether to use stated or inferred ANA, or indeed some mix of the two, the inference of ANA using an analytical approach has strong appeal. Extra information does not need to be collected, and issues of endogeneity do not need to be addressed with respect to the stated ANA responses (e.g., Hensher, 2008). However, each of the analytical methods in the literature may be inappropriate in certain circumstances. The use of conditional parameter estimates by Hess and Hensher (2010) relied on sequential estimation, as well as a threshold that was selected somewhat arbitrarily by the analyst. Further, Mariel et al. (2011) showed that the threshold itself varies with the true rate of ANA. Censored random parameter (RP) distributions (Train and Sonnier, 2005) can capture ANA, but may confound ANA with preference heterogeneity, since the same structural parameters are controlling both the continuous and discrete components of the distribution. The use of the conventional LC model (Swait and Adamowicz, 2001) may require many classes if ANA behaviour is complex, and may not be effective if ANA for one attribute varies independently of the tastes and ANA of the other attributes. Hess and Rose (2007) employed a form of LC model, whereby constraints were imposed on each class, reflecting a particular combination of choice heuristics. This approach has been widely adopted in the literature, however as nonattendance is modelled for an increasing number of attributes, or as other heuristics are added, the number of classes increases rapidly. Since the number of parameters will increase exponentially with respect to the number of attributes, estimation becomes increasingly challenging. Gilbride et al. (2006) used a modification of the variable selection model (George and McCulloch, 1993) to identify ANA within the Bayesian framework, where this framework may preclude widespread adoption[1]. There is scope for an analytical method which does not rely on selection of a threshold, does not require an unfeasibly large number of parameters for typical choice scenarios, and which can be estimated in the classical framework.

The LC approach of Hess and Rose (2007) discussed above handles ANA by either omitting an attribute's parameter from the utility expression of a class, if that class is associated with the nonattendance of the attribute in question, or including the parameter if the class represents attendance. However, typically, the same parameter enters all classes that represent attendance to the attribute, as distinct from a conventional LC model, where unique parameters are estimated for each class. In all applications of the model, this parameter is

[1] Classical econometric methods are widely used, and the learning curve for Bayesian methods is steep (Train, 2009).

a point estimate, whereby a single coefficient enters the representative utility. Consequently, the only preference heterogeneity that can be captured is bimodal in nature. Either the coefficient is zero, or the value of the parameter specified. This thesis will show that such a specification, in the presence of preference heterogeneity of those who attend to an attribute, may confound attribute attendance with preference heterogeneity. This in turn will bias both the measure of ANA, and the mean preference of attenders. This is a motivation both to extend the LC approach to overcome this problem, and to ensure that any proposed new analytical method does not likewise introduce bias. The use of random parameters within the LC approach may overcome the above problem, and is further motivated by the paucity of studies that combine discrete and continuous mixed logit models (Lenk and DeSarbo, 2000; Bujosa et al., 2010; Hensher et al., 2012b; Hess et al., 2011).

1.3 Contribution to the literature

This thesis provides a systematic examination of the impact of ANA on parameter estimates, when ANA is not accounted for. Extensive use is made of simulations, to overcome the problem of not knowing the true parameters and rates of ANA in empirical datasets. Three dimensions are varied in the simulations: the ANA rate, the distribution of preference heterogeneity of attribute attenders, and the measure of dispersion of the distribution. The models estimated on the simulated datasets are evaluated in terms of the resultant bias in the moments of the random parameters. For applicable distributions, the percent of coefficients that are implausibly signed is also calculated. This component of the thesis contributes to the literature by

(a) identifying the systematic nature, and extent, of the bias of all structural parameters, when RPL models are estimated in the presence of ANA;

(b) determining the sensitivity of the percentage of implausibly signed coefficients to changes in the true ANA rates and the magnitude of the preference heterogeneity; and

(c) evaluating the performance of the LC approach to handling ANA, by using simulations where the true values are known.

In addition to gaining a more complete understanding of the consequences of not handling ANA, this thesis introduces a model that can analytically determine the rate of ANA. The model possesses a number of important features. While each of these features are shared with at least one existing method in the literature, no existing method possesses them all. The features include:

Analytical determination of ANA, allowing estimation that is not reliant on stated ANA, and avoiding problems of endogeneity;

Covariates, whereby additional information (including, potentially, stated ANA) may be predictors for the nonattendance of each attribute;

No arbitrary thresholds, or any need for ex-ante decisions by the analyst on values or model specifications that lack a behavioural interpretation;

Simultaneous estimation, of ANA and other model outputs;

Taste heterogeneity, beyond bimodal attendance/nonattendance. A point mass is estimated for nonattendance, while a set of structural parameters describe taste heterogeneity over those respondents that attend;

Parsimony, specifically, the ability to limit the number of parameters controlling ANA, as the number of attributes tested for ANA increases;

Classical estimation, rather than in the Bayesian framework.

The performance of the new model is tested on simulated datasets, allowing both the outright performance of the model to be tested (against known values), as well as the relative performance (against the existing methodologies). The model is also estimated on an empirical dataset, and comparisons are made against the stated ANA rates, and some of the other models that can capture ANA. A specific contribution, enabled by the properties of the model, is the examination of whether covariates can be identified that influence the rate of ANA. In particular, the use of stated ANA as a covariate provides further evidence in the unresolved debate concerning the reliability of stated ANA. Some studies have shown that covariates might add explanatory power, but have only done so in a framework that does not scale well as the number of attributes that may be ignored increases (e.g., Hess and Rose, 2007). The thesis also examines whether the best distributional assumptions of the random parameters changes once ANA has been accounted for. This is plausible, due to the way in which the continuous distribution can approximate ANA with the RPL model. The proposed model is parsimonious in nature, but the most parsimonious specification relies upon an assumption that ANA is independent across all attributes. To maximise parsimony, without necessarily relying on this assumption for all attributes, the model is specified in a generalised way, such that parsimony can be compromised as required, where the independence assumption does not hold.

1.4 Outline of the thesis

In very broad terms, this thesis critically evaluates the literature on ANA (Chapter 2), proposes a new methodology for handling ANA that overcomes the shortcomings of existing methods (Chapter 3), investigates further the influence of ANA when it is unchecked, and the properties of two existing methodologies (Chapter 4), tests the new methodology with simulated and empirical data (Chapters 5 and 6), and discusses the findings and outlines a future research agenda (Chapter 7). Each of these chapters are now considered in detail.

Chapter 2 begins with a formalisation of the widely employed discrete choice models that will serve either as a reference point in this thesis, or as the building blocks of the model that is introduced. The chapter then reviews the extensive literature on ANA. Attribute nonattendance is defined, and an acknowledgment is made of the relevant broader literature, which spans econometrics, psychology and various applied fields such as transportation and environmental economics. Reasons why ANA occurs are expounded, and the literature is drawn upon to argue why the study of ANA is important, and why it may need to be handled in discrete choice models. Each of the various methods for handling ANA are then critically evaluated in turn. Next, the review steps back from the specific methodologies, and considers a number of overarching themes that have been identified to varying degrees in the literature. Broadly, these themes concern the nature of ANA behaviour, the impact of ANA on models that are poorly suited to handling it, and some fundamental approaches to handling ANA behaviour in discrete choice models. Drawing upon this review, the weaknesses of current ANA methodologies are identified, an alternative methodology for accommodating ANA is proposed in general terms, and this methodology is framed in terms of how it improves upon the alternatives, and contributes to the literature.

Chapter 3 formalises the model that was proposed in general terms in Chapter 2. First, a model is introduced that generalises two variants of the LC model that have been employed to handle ANA in the literature. At one extreme, maximum parsimony is achieved by assuming that ANA is independent across all attributes. At the other, full correlation in ANA can be achieved, albeit at a parametric cost. The generalised approach allows for many intermediate points, each with a partial assumption of independence of ANA. It is suggested that the model be referred to as the attribute nonattendance (ANA) model. Next, the model is extended to include random parameters, and allow for a continuous specification of preference heterogeneity, conditional on the attribute being attended to. This is referred to as the random parameters attribute nonattendance (RPANA) model. Issues with the estimation and identification of the RPANA mode are noted, and solutions provided. Finally, the chapter summaries some key papers employing ANA models in the literature, and positions the RPANA model against these.

Chapter 4 employs simulations to systematically quantify the nature and severity of the biasing influences of ANA on the sensitivities recovered by the RPL model. The use of simulated data overcomes the problem with empirical studies of not knowing the true sensitivities or ANA rates. Three dimensions are varied, allowing interactions between them to be studied: the ANA rate, the true distribution of preferences, and the extent of preference heterogeneity. After outlining the aims and motivations of the chapter, the methodology is detailed. This includes the data generation process, the discrete choice models estimated, and how the results are analysed. Findings are then presented for five different taste distributions. The latent class based ANA model is also estimated on the simulated datasets, to gauge the performance of the model in terms of the accuracy of recovery of the true sensitivities and ANA rates. From these simulations, the value of the RPANA model is demonstrated. Finally, the findings are summarised and discussed.

Chapter 5 uses the same simulated datasets to evaluate the performance of the RPANA model. It is shown that the RPANA model recovers the true values with a high degree of accuracy.

Chapter 6 evaluates the performance of the RPANA model on an empirical dataset. The application area, air travel behaviour, is first outlined, followed by details of the specific empirical study. A series of models are estimated that serve as reference points, including the MNL, RPL and ANA models. Then, various specifications of the RPANA model are compared, both to each other, and to the reference models, to gain a nuanced understanding of the strengths and weaknesses of the RPANA model.

Chapter 7 critically evaluates the performance of the RPANA model, and compares the model to a number of other approaches. Some impediments to its adoption are also discussed. Informed by the findings of the thesis, the chapter revisits and introduces some behavioural and econometric issues associated with ANA. A future research agenda is suggested, before the thesis is concluded.

Chapter 2

Literature review

Before reviewing the literature on ANA, it is first necessary to precisely outline the family of discrete choice models within which ANA has been explored. Thus, Section 2.1 details the MNL model, the RPL model, and the LC model. These models will form the building blocks upon which the models proposed in this thesis rely. They will also serve as behavioural and econometric reference points for the proposed models.

Section 2.2 defines ANA, and evaluates the many reasons that have been proposed for why ANA occurs. Attribute nonattendance is also recognised as being positioned within a broader literature concerned with choice heuristics and decision rules. Section 2.2.1 argues that the study of ANA is important, and that ANA should be identified and handled in discrete choice models. The arguments draw upon evidence from the literature.

Each of the existing methods that have been employed in the literature to accommodate ANA are outlined in Section 2.2.2. Variants on each method are detailed, findings from the literature are summarised, and the methods are critically appraised. Section 2.2.3 steps back from the specific methods, and considers some of the key issues in the ANA literature. Some of these issues are well recognised and extensively discussed in the literature. Examples include whether to rely on revelations of ANA behaviour by respondents in SC studies, or determine ANA analytically; whether individuals are consistent in their ANA behaviour across multiple choice tasks, or whether it may vary from one choice task to the next, and the methodological implications of each of these alternatives; and the impact on WTP of ANA. Other issues have been addressed by the literature, but less extensively. Examples include heterogeneity in ANA behaviour across respondents, systematic investigations using simulations of the impact of ANA, and the interrelation of ANA and implausibly signed RPL coefficients. In Section 2.3, gaps in the literature are explored, the broad research agenda of this thesis is outlined, and its contribution is noted.

2.1 Discrete choice modelling methodology

2.1.1 Multinomial logit model

The MNL model (McFadden, 1974) is the most widely used discrete choice model. Even where more advanced models are estimated, the MNL model serves as a useful reference point. Since the log-likelihood of the model is globally concave, so long as the utility is linear-in-parameters, a globally optimal solution is guaranteed. One assumption of the MNL model that may be overly restrictive, particularly in the context of this thesis, is that of homogeneity of preferences. That is, unless interactions are specified to capture systematic preference heterogeneity, the model assumes that all individuals have the same tastes for the attributes of the choice alternatives. The LC and RPL models will relax this assumption.

Consider first the total utility of alternative i on choice occasion t for respondent or individual n, U_{nit}, which is composed of the representative utility V_{nit}, and the unobserved component of utility, ϵ_{nit}. The representative component is associated with a vector of observed variables and alternative specific constants (ASCs), x_{nit}. The utility associated with these variables is estimated with a vector of taste coefficients β, such that the representative utility is $V_{nit} = \beta x_{nit}$. For the MNL model, the probability that alternative i will be chosen is

$$P_{nit} = \frac{e^{\beta x_{nit}}}{\sum_{j=1}^{J} e^{\beta x_{njt}}}. \tag{2.1}$$

To estimate the taste coefficients, we may employ maximum-likelihood estimation. Consider a sample of N respondents, each completing T choice tasks, with each choice task containing J alternatives. We observe a vector y of binary choice outcomes for every alternative in every choice task, such that y_{njt} equals one if the alternative is chosen, or zero otherwise. The log-likelihood function is then

$$\mathrm{LL}(\beta) = \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} y_{njt} \ln P_{njt}, \tag{2.2}$$

which we maximise to obtain the maximum-likelihood values of the taste coefficients β.

2.1.2 Latent class model

The LC discrete choice model (Kamakura and Russell, 1989) relaxes the assumption of preference homogeneity. Multiple classes of taste coefficients are estimated, as are the probabilities that each respondent resides in each class. The analyst chooses the number of classes prior to model estimation. Various numbers of classes may be tested, and model fit compared using a measure such as the Akaike Information Criterion (AIC) or Consistent Akaike Information Criterion. The taste coefficients in each class are typically freely estimated, but constraints may be imposed across classes (e.g., Hess and Rose, 2007).

Denote the set of all classes as M. We need to estimate the probability of a respondent n residing in each class. This is achieved with an MNL model, specified as

$$P_{nm} = \frac{e^{(\gamma_m + \theta_{nm} z_n)}}{\sum_{d \in M} e^{(\gamma_d + \theta_{nd} z_n)}}. \tag{2.3}$$

Here, a parameter γ_m serves as a constant term, capturing the assignment to class m that cannot be explained by other factors. A vector of parameters, θ_{nm}, captures socio-demographic and other influences on the assignment of respondent n to class m, where z_n is a vector of those influences. Alternatively, γ_m could be an element of θ_{nm}, but it is denoted separately in recognition of the frequent estimation of a constants only class assignment model. To ensure identification, γ_m and θ_{nd} are constrained to zero for one class.

Next, consider the probability of alternative i being chosen on choice occasion t, conditional on respondent n being assigned to class m. This probability is typically also generated with an MNL model.

$$P_{nit|m} = \frac{e^{\beta_m x_{nit}}}{\sum_{j=1}^{J} e^{\beta_m x_{njt}}}. \tag{2.4}$$

Note that β_m is the vector of taste coefficients associated with class m.

An important advantage of the LC model is that, in addition to multiple taste coefficients being estimated for each observed variable, the probability for each respondent of being assigned to each class, and hence each taste coefficient across classes for any observed variable, can be assumed to be constant over choice occasions. That is, by assuming tastes to be stable across choice occasions for each respondent, the panel nature of the data can be exploited. Consider a sequence of choices of alternatives over T choice occasions, $\vec{i} = \{i_1, \ldots, i_T\}$. Assuming that the unobserved component of utility is independently and identically extreme value type 1 distributed over alternatives, respondents, *and* time, the probability of a sequence of choices of alternatives, conditional on assignment to class m, is

$$P_{n\vec{i}|m} = \prod_{t=1}^{T} \left[\frac{e^{\beta_m x_{ni_t t}}}{\sum_{j}^{J} e^{\beta_m x_{njt}}} \right]. \tag{2.5}$$

The unconditional probability of a sequence of choices, $\vec{i}$, for respondent n is obtained by taking the product of two probabilities: assignment to class m, and the probability of the sequence of choices, conditional on assignment to that class; then integrating over all classes. This can be expressed as

$$P_{n\vec{i}} = \sum_{m \in M} P_{nm} P_{n\vec{i}|m}. \tag{2.6}$$

Substituting in Equations 2.3 and 2.5, Equation 2.6 becomes

$$P_{n\vec{i}} = \sum_{m \in M} \frac{e^{(\gamma_m + \theta_{nm} z_n)}}{\sum_{d \in M} e^{(\gamma_d + \theta_{nd} z_n)}} \prod_{t=1}^{T} \left[\frac{e^{\beta_m x_{ni_t t}}}{\sum_j^J e^{\beta_m x_{njt}}} \right]. \tag{2.7}$$

The log-likelihood function is

$$\text{LL}(\beta) = \sum_{n=1}^{N} \ln P_{n\vec{i}}. \tag{2.8}$$

Also, the posterior class assignment probabilities can be computed for each class m, conditional on the sequence of choices $\vec{i}$ by respondent n, as

$$P_{m|n\vec{i}} = \frac{P_{nm} P_{n\vec{i}|m}}{\sum_{d \in M} P_{nd} P_{n\vec{i}|d}}, \tag{2.9}$$

(see Greene and Hensher, 2003). Further, the individual specific parameter estimates can be computed as

$$\beta_{n\vec{i}} = \sum_{m \in M} P_{m|n\vec{i}} \beta_m. \tag{2.10}$$

2.1.3 Random parameters logit model

An alternative way to capture preference heterogeneity is with the RPL model (see Hensher and Greene, 2003; Train, 2009). Instead of a discrete mixture of taste coefficients, as in the LC model, the taste coefficients β vary over decision makers with continuous density $f(\beta)$. A distribution is specified for each taste coefficient, and the moments of these distributions are estimated with structural parameters. Alternatively, a joint distribution, such as the multivariate normal, can be specified. As with the LC model, the panel nature of the data can be exploited, with the coefficients assumed to remain constant over all choice tasks completed by a respondent. The unconditional probability of a sequence of choices $\vec{i}$, for respondent n, is obtained by integrating over $f(\beta)$.

$$P_{n\vec{i}} = \int \prod_{t=1}^{T} \left[\frac{e^{\beta x_{ni_t t}}}{\sum_j^J e^{\beta x_{njt}}} \right] f(\beta) d\beta. \tag{2.11}$$

The posterior parameter estimates are not detailed here, as while they are employed in some parts of the ANA literature, they are not employed in this thesis. The reader is referred to Hensher and Greene (2003) or Train (2009) for details.

When specifying a RPL model, the analyst needs to make several decisions. Crucially, the parameter distributions must be decided upon. The analyst has a large number of distributions that can be employed, including, but not limited to, the normal, triangular, uniform, lognormal, Rayleigh, and Johnson's S_B, as well as constrained and censored versions of some of these distributions. The choice of distribution is likely to be an empirical issue, although a distribution that is constrained in sign to be all positive or all negative will have behavioural appeal in many circumstances. Indeed, the tension between econometric and behavioural appeal will be a theme of this thesis. The number of draws used to simulate the probabilities must also be decided upon, where Walker (2002) warns of the danger of employing too few draws.

2.2 Attribute nonattendance (ANA) in discrete choice models

A typical assumption of discrete choice models is that all attributes are considered by all individuals, to some degree, and that all of these attributes are either traded off in a compensatory manner, or have no influence on choice for all individuals. Challenging this assumption is the notion of attribute nonattendance, in which an attribute of a choice alternative influences the choices of only a subset of individuals. For the remaining individuals, even if the attribute is observed prior to choice, it has no influence when the choice is made. Another way of considering ANA is that there may be differences between individuals in terms of which attributes have any influence on their choice. This has widely been referred to in the literature as the ignoring of attribute (e.g., Hensher et al., 2005), and, as used in this thesis, attribute nonattendance (Scarpa et al., 2009). Crucial to an understanding of ANA is a consideration of the reasons why it might occur. These fall into three broad categories, which will be detailed in turn. First, ANA might reflect a genuine disinterest in the attribute by the individual, and consequently be somewhat immutable. Second, and in contrast, ANA might be a phenomenon that is context dependent; employed in response to some property of the choice task at hand, or some change in the individual, for example, fatigue. Third, ANA might reflect strategic behaviour, where this is especially likely in an SC experiment focused on non-market valuation or public policy choices.

The simplest explanation for ANA is that some individuals have a genuine disinterest in the state of an attribute, and thus ANA is purely a form of preference heterogeneity, albeit one that is extreme (Balcombe et al., 2011). Gilbride et al. (2006, p.420) suggest three possible antecedents for this form of preference heterogeneity: "variation in motivations, expertise, and/or perception". Whilst these components can change over time, as can preferences overall, it is reasonable to assume that they will remain constant for any one individual, in the short term. This is in contrast to an extensive collection of ANA causal factors, in which the unifying trait is that ANA behaviour is prone to vary, in the short term, across choice occasions. Drawing a distinction between these two broad explanations for ANA does not mean that they are mutually exclusive. It is likely that preference heterogeneity will have an influence even when ANA behaviour varies across choice occasions, moderating the propensity to attend. What is distinct about the extreme preference heterogeneity explanation is that an individual may be indifferent to an attribute under any circumstances. They simply may not care.

Attribute nonattendance may be employed by the individual as a mechanism for coping with choice complexity. The concept of bounded rationality (Simon, 1972) recognises that decision makers possess limited cognitive capabilities, and exhibit limited motivation to process information. DeShazo and Fermo (2004) contrast two important forms of bounded rationality. Under the passive bounded rationality model, full attention is retained as the complexity of the choice task increases, with the cognitive limits of the individual resulting in a greater propensity to make errors as the information is processed. This has been detected in econometric discrete choice models through an increase in error variance (Arentze et al., 2003; Caussade et al., 2005). Under the rationally-adaptive model, the individual responds to complexity by selectively attending to information. The individual is aware that there are limitations both to their capabilities, and to the benefits of attention. Further, a cost is incurred as an attribute is attended to, and its information processed and integrated as a part of the overall decision making process. The propensity to attend may vary both across individuals, due to differing capabilities and motivations (Hensher et al., 2005), and across choice tasks, if the complexities of the choice tasks differ (Hensher, 2006a). A key feature of the rationally-adaptive model is that the individual makes decisions as they process the information, prior to choosing one of the choice alternatives. Precisely what those decision rules are is not prescribed by the model, and needs to be identified (DeShazo and Fermo, 2004).

DeShazo and Fermo (2004) found evidence for the rationally adaptive model. They systematically varied the complexity of their designs via the number of alternatives and attributes, as well as the correlation structure of the information. Complexity was found to have an impact on both the error variance and the attendance to the attributes, although accounting for ANA reduced the impact on the error variance. This suggests that choice complexity influences the propensity to attend, as well as choice consistency. Hensher (2006a) also investigated the impact of choice task complexity on the propensity to attend to attributes. Within the choice tasks of an SC study, he varied, across respondents, the range of the attribute levels relative to a reference alternative that was a recent car trip, and the number of alternatives, attributes, and attributes levels. He found that the aggregate ANA rate increased as the number of attribute levels increased, the difference between attribute levels decreased, the number of alternatives decreased, and the range of the attribute levels relative to the reference alternative decreased. With the exception of the third finding, all of these outcomes suggest that increasing the cognitive load leads to rationally-adaptive outcomes. Rose et al. (2009), however, performed a cross cultural study, and found that some of the differences in processing may be culturally specific. Alemu et al. (2011) asked respondents of an SC study not just which attributes they ignored, but why they did so. One of the categorical responses available was "It made it easier to choose between the alternatives" (p.11). This reason is consistent with the rationally-adaptive model, as it implies that the individual is reducing the amount of information, to reduce their cognitive load. Of the five categories presented, this response was the most frequently chosen, accounting for an average of 48 percent of responses that an attribute was ignored, lending credence to the rationally-adaptive model.

In the context of SC studies, one possible strategy for handling complexity and reducing the incidence of ANA is for the analyst to make the choice task simpler. Despite the widespread use of choice tasks with low dimensions (e.g., Fosgerau, 2006), doing so to alleviate concerns about complexity may not be advisable. Gilbride et al. (2006) warned that simplified choice tasks may not reflect market choices. Collins et al. (2012) presented survey respondents with a choice task that contained a large number of alternatives. Respondents could utilise structured search mechanisms to remove and reorder information. Error variance was *lower* than a conventional choice task that contained just three alternatives.

Hensher (2006b) suggested that more information in a choice task need not equate to an increase in complexity, as a simplified choice task might not contain an amount of information deemed appropriate by an individual. What matters is not information quantity, but *relevance* (Hensher, 2006a). The problem with an analyst making decisions about what information should be retained in an SC task, so as to ensure simplicity, is that the analyst's prior expectations as to what is relevant may not align with what matters to the respondent. So long as incomplete attention can be handled adequately, providing a richer set of *potential influences* to each respondent, or considering more attributes in a revealed preference study, may lead to more consistent choice by the decision maker, and deeper behavioural insights by the analyst.

Attributes might not be attended to due to the cost of acquiring information about the attribute. In an SC choice task, all information included by the analyst is readily available to the respondent. While processing this information will involve cognitive effort, and cost-benefit decisions may result in ANA, the information *acquisition* is close to cost free. In contrast, real choices made in the market are frequently made over choice alternatives for which much of the attribute information may not be readily available, for a variety of reasons. Acquiring the information will incur some cost, and so there may be varying degrees of information acquisition, and hence attendance, across individuals and attributes.

Rose et al. (2012a) illustrate with the example of the torque of a car's engine. Of this attribute, an individual may have acquired perfect information, limited information, or may not have acquired any information at all. Even if information is acquired, the individual may not well understand what it means, and may not even know if more of the attribute is better or worse. At the extreme, the individual may not even know that the attribute exists[1]. Extending upon the points of Rose et al. (2012a), the individual might know that some attribute information exists, but not know how to reveal it, and so consider it not acquirable[2]. If the individual knows little about the attribute, then it is reasonable to expect that it may be somewhat 'expendable', and might be an early candidate for nonattendance if simplification of the choice task is sought by the individual. This would likely be particularly so for attributes for which the respondent does not know whether more is better.

Some aspects of ANA bear resemblance to the consideration set literature, which recognises that only a subset of available *alternatives* might be considered. While one interpretation of consideration sets is that they result from constraints imposed upon individuals (e.g., Swait and Ben-Akiva, 1987), they are more frequently explained with respect to search costs and marginal benefits resulting from the evaluation of the next choice alternative (Hauser and Wernerfelt, 1990; Roberts and Lattin, 1991). This is an alternative-based means of processing information (Payne et al., 1988), whereby the alternatives are considered sequentially, although consideration set generation has also been explained as a process of attribute com-

[1]Even more extremely, it may be a Rumsfeldian 'unknown unknown'.

[2]Consider for example the placement of an in-flight entertainment unit box under some airline seats, which reduces legroom. The individual may know this is a potential problem, but no have any idea how to find out if a particular seat is impacted. Websites such as www.seatguru.com provide this information to those 'in the know'.

parison (e.g., Gensch, 1987). Attribute attendance could be considered as a consideration set of attributes, with the size of the set being dictated by costs and benefits in a rationally-adaptive manner, as with most models of (alternative) consideration set formation.

Attribute nonattendance is just one of many mechanisms that an individual might employ as they process information, prior to making their choice. These mechanisms have been described variously as heuristics, decision rules, attribute processing strategies, and information processing strategies. Whilst the psychology literature has been the traditional conduit for research into these mechanisms, there has also been a growing research effort that aims to incorporate heuristics and decision rules into econometric models of choice. Across a range of literatures, examples include, but are by no means limited to:

Elimination by aspects: sequential elimination of alternatives based on specific attribute level criteria, until only one alternative remains (e.g., Tversky, 1972; Batley and Daly, 2006);

Lexicographic choice: persistent choice of the alternative that performs best on a particular attribute (e.g., Saelensminde, 2006; Hess et al., 2010);

The majority of confirming dimensions heuristic: choice of the alternative with the greatest number of superior attribute levels (e.g., Russo and Dosher, 1983; Hensher and Collins, 2011);

Satisficing: choice is made when utility exceeds an aspiration level, rather than when it is maximised (e.g., Simon, 1956);

Regret minimisation: choice is driven by avoidance of negative emotions (e.g., Chorus et al., 2008);

Non-trading: persistent choice of the same (usually labelled) alternative (e.g., Hess et al., 2010);

Referencing and prospect theory: choice alternatives are evaluated in terms of gains and losses, relative to some reference point (e.g., Kahneman and Tversky, 1979; Hess et al., 2008);

Aggregating common-metric attributes: the partitioning of attributes with the same unit (e.g., dollars) is either retained, or the attributes are aggregated (e.g., Layton and Hensher, 2010; Hensher and Greene, 2010);

Parameter transfer rules: marginal utilities for common-metric attributes either remain distinct, or assume the marginal utility of the attribute with the dominating level (e.g., Hensher and Layton, 2010);

Thresholds on attribute levels: choice alternatives are penalised if particular attribute levels are exceeded (e.g., Swait, 2001).

A trait common to many of these heuristics and decision rules is that the individual is likely to use only a subset of the information, and thus not attend to some of the attributes. Consequently, ANA (or conversely, selective attendance) can be considered as a building block upon which many of the heuristics depend. This is both a blessing and a curse: ANA is highly relevant, but there is a risk that ANA might be confounded with other, potentially more complex heuristics.

Some fundamental differences exist between the way that choice heuristics and decision rules are captured and handled in the psychology literature, and in econometric choice models. Within the psychology literature, process tracing methods are often employed to observe how an individual processes information prior to choice. For example, the information board (Ford et al., 1989) presents a matrix of alternative and attributes, but initially hides all attribute levels, requiring mouse clicks to reveal each attribute level (i.e., alternative-attribute combination). This technique has helped uncover and validate a wide variety of heuristics. It allows the uncovered heuristics to vary across individuals, between choice tasks for each individual, or even within choice tasks. However, the experimental conditions used in these studies are typically highly artificial, raising questions about the validity of the findings beyond the laboratory. For example, in many of these experiments, only one attribute can be viewed at a time. Further, such experimental conditions may induce demand effects, resulting in processes being employed that would not be used in a more natural environment.

Econometric methods are different in a number of key respects. The pre-choice information processing of an individual is typically not observed. Whilst the models may seek to determine differences between individuals, for example in terms of preferences, or error variance, they are typically aggregate models that share information across individuals (Gilbride et al., 2006). Further, the models rely on a number of assumptions. Conventionally, it is assumed that the decision maker is indefatigable, attends to and incorporates all information, and integrates this information in a compensatory manner (Swait and Adamowicz, 2001).

A number of challenges are faced when heuristics and decision rules are embedded in econometric models of discrete choice. The heuristic may violate some assumption of the basic model, necessitating some change to the econometric specification to overcome the violation. Handling multiple heuristics may be difficult, as it may lead to confounding between the heuristics (Hess et al., 2012), and may be parametrically expensive. Nonetheless, some progress has been made (Swait and Adamowicz, 2001; Hensher and Greene, 2010; Hensher and Collins, 2011; Leong and Hensher, 2011; Hess et al., 2012). Ultimately though, there will be limits to the amount of complexity that can be introduced into the model, and supported by the data.

Given the wide spectrum of heuristics and decision rules, the researcher is faced with a decision as to where to focus their research effort (Swait and Adamowicz, 2001). In the context of econometric models, ease of estimation is appealing, as a tractable model is more likely to be employed in practice. Scalability across multiple heuristics is also an advantage. Ultimately though, what is important is the impact that accommodating the heuristic or decision rule has on model outputs and model fit, and the broader implications in whatever field the choice model is being applied.

In the context of SC studies, ANA may be a consequence of the experimental design employed. The levels of the attributes may be of inappropriate range, or may not matter to the respondent (Hensher et al., 2012a). Alemu et al. (2011) found that of the stated responses indicating ANA, for an average of 11 percent the respondent conceded that "The levels for the attribute were unrealistically high/low" (p.11). Hensher et al. (2012a) speculated that an attribute threshold may need to be reached before attendance is given, while Campbell et al. (2012) provided empirical evidence to this effect. In addition to the attribute ranges, ANA might result from behaviourally questionable tradeoffs in the choice tasks (Hensher et al., 2012a).

If the design is realistic and well suited to the problem at hand, then ANA that is induced by attribute ranges or tradeoffs may be acceptable. For example, toll costs may be ignored by some high income individuals for any range that would realistically be implemented. So long as ANA is handled by the model, a design that allows these respondents to still ignore the attribute is not problematic. Indeed, it may be preferable to a design that includes extremely high tolls, as this may undermine the plausibility and credibility of the study. Alternatively, respondent specific attribute ranges could be generated through pivoting (Rose et al., 2012a). Pivoting the attribute levels around some reference point that is meaningful to the respondent may lead to attribute levels that are more consistent with their experiences, and so reduce ANA that is induced by inappropriate attribute level ranges. Nonetheless, an appropriate range of attribute levels must be selected, and the pivoting may not reduce ANA if the respondent is insensitive to all levels near their recent experience. Rose et al. (2012a) noted that techniques for generating experimental designs have focused on statistical efficiency and reduced cognitive effort, but not differences in the way that people process information. How this could be achieved remains an area for future research.

Attribute nonattendance may also result from strategic behaviour, or occur because an individual may not believe that certain attributes should be traded. Non-market valuation studies are particularly likely to be suspectable to such problems. Scarpa et al. (2009) found very high nonattendance rates to cost in the public policy domain, and speculated that respondents might have had difficulty in trading off aspects of rural landscape with money. It is plausible though that those respondents that ignored cost did so in *protest* against having to place a monetary value on the landscape; enacting such a protest would be very *easy*. Alemu et al. (2011) found that of those responses to indicate that an attribute was ignored in a choice task, 13 percent were because the respondent did not "think that this attribute should be weighed against the others" (p.11).

Clearly, there are many reasons why ANA might occur. However, before the complexity of accommodating ANA into econometric discrete choice models can be entertained, it is worth considering why this is an important endeavour. What follows is a summary of why accommodating ANA is important.

2.2.1 Motivations for accommodating attribute nonattendance in discrete choice models

There are many compelling reasons for identifying and accommodating ANA in choice models. Widely acknowledged in the literature is the detrimental impact of bias in parameter estimates and WTP measures, resulting from ANA. Latent class and RPL models are likely to assign nonzero marginal utilities when they should be zero, resulting in this bias (Rose et al., 2005, 2012a). The impact of such bias will depend on how the model outputs are applied, but as an example, biases in values of travel time savings (VTTS) may lead to large differences in travel time benefits (Hensher et al., 2007), and impact on economic appraisal and demand forecasting (Hensher and Greene, 2010). Section 2.2.3 provides an extensive discussion of the impact of ANA on WTP.

Accommodating ANA may uncover a picture of preference heterogeneity that is both more nuanced, and more plausible. Sensitivities which may otherwise be insignificant may be revealed as significant for the subsample that attends to an attribute (Rose et al., 2005, see Section 2.2.2). Handling ANA may reduce the incidence of implausibly signed coefficients in RPL models (Hensher, 2007, see Section 2.2.3). Failing to handle ANA may result in overinflated measures of preference heterogeneity (Hensher, 2007; Puckett and Hensher, 2008; Campbell and Lorimer, 2009; Hess and Hensher, 2010).

Attribute nonattendance may be of intrinsic interest to the analyst. For example, managers may wish to know who is not attending to certain product features, so that they can target a marketing campaign to such individuals (Rose et al., 2012a). This would be particularly useful if the attribute ignored was one on which the company had a competitive advantage. The idea here is to promote sales through an expansion of the individuals' 'consideration set' of attributes, rather than a more direct expansion of their consideration set of alternatives (Shocker et al., 1991; Roberts and Lattin, 1997).

Attribute nonattendance may also facilitate a critical appraisal of the experimental designs of choice experiments. Numerous authors have noted that ANA might be induced by an experimental design with inappropriate levels and tradeoffs (Hensher et al., 2005; Scarpa et al., 2009; Rose et al., 2012a; Hensher et al., 2012a). Higher than anticipated rates of ANA may alert the analyst to problems with the design, and allow them to revise the design before the main wave of a survey. Whilst accommodating ANA might reduce the bias in already collected data, it would be better to identify the problem early and collect superior data.

Broadly, accommodating ANA helps instill confidence that latent preferences are being captured adequately by the choice model. This is achieved by reducing bias in taste coefficients and WTP measures, reducing the incidence of implausibly signed coefficients, reducing excess preference heterogeneity, and either accommodating a questionable experimental design, or flagging such a design for improvement. If ANA can be inferred analytically, then these outcomes can be achieved merely through an appropriate model specification, without additional data, furthering the appeal.

2.2.2 Existing methodologies

A range of methods have been proposed for detecting and accommodating ANA. These methods fall into two broad camps: stated ANA, wherein respondents are asked which attributes they attended to, and the analyst uses this information in some way; and inferred ANA, wherein the choice model is specified in such a way that ANA can be retrieved analytically. What follows is a detailed overview of how the literature has employed these two methods, starting with stated ANA, and progressing through the various analytical methods.

Stated attribute nonattendance

Stated ANA relies upon questions posed to individuals about which attributes, if any, were ignored when they made their choice(s). The first two papers that we know of that utilised stated ANA were Hensher et al. (2005) and Rose et al. (2005), who obtained the stated ANA responses after all of the choice tasks were presented to the respondent. Stated ANA was accommodated econometrically by setting the marginal utilities to zero, in those instances in which an attribute was stated as being ignored. A drop in VTTS was observed in both studies, once ANA was accounted for. Neither drew broader conclusions about the direction of this change. In particular, Rose et al. (2005) stated that failing to account for ANA will *likely* lead to some degree of bias, but that this could either be upwards or downwards. This issue will be discussed extensively in Section 2.2.3.

In addition to changes in VTTS, Rose et al. (2005) found that an attribute that was insignificant in the naïve model became highly significant and of the correct sign once ANA was accounted for. They noted that ANA could be used as a 'segmentation criterion', and that sensitivity to an attribute may not be captured if the segment that attends to the attribute is small, and ANA is not adequately handled. Thus, handling ANA can assist in uncovering preference heterogeneity that would otherwise be masked by an aggregation bias.

Hensher et al. (2005) cautioned that unconscious preferences might play a role in decision making, and questioned whether stated ANA can be sustained under other tests. For example, Alemu et al. (2011) asked a multitude of questions and related them to ANA, while much of the literature forgoes stated ANA entirely, and relies on analytically inferred ANA. These analytical methods will be detailed in subsequent sections, and the stated and analytical approaches will be compared and discussed in detail in Section 2.2.3.

Rose et al. (2005) and Hensher et al. (2005) asked respondents just once whether they ignored any attributes, after all choice tasks were completed, thereby capturing what is frequently referred to as serial ANA. Others have obtained these responses after each choice task, and then conditioned the marginal utilities at the level of the choice task, rather than the respondent (Puckett and Hensher, 2008, 2009; Meyerhoff and Liebe, 2009; Scarpa et al., 2010). Section 2.2.3 outlines the findings from this second approach, known as choice task ANA, and discusses a range of related issues. These include the burden of additional data collection, evidence for within respondent differences in ANA across choice tasks, the consequences of any differences for analytical ANA methods, and whether inconsistent ANA behaviour is reflective of choice task influences on ANA.

An alternative approach to determining whether an individual ignored an attribute in an SC experiment is to allow them to control whether or not an attribute is actually displayed at all. If the level of an attribute is never revealed, then it is reasonable to assume that the attribute was ignored. Kaye-Blake et al. (2009) utilised the information display board approach (Ford et al., 1989). A conventional table format was employed to present the three choice alternatives and their respective attribute levels. Initially, no attribute levels were shown, and the respondent had to click on each alternative-attribute combination to reveal the corresponding level. Once exposed, the attribute level would remain exposed until the next choice task, whereupon all levels would once again be obscured. Thus, any level not revealed could be considered as ignored. This approach revealed ANA at both the choice task and alternative level. Whereas Puckett and Hensher (2009) found little difference in ANA across the alternatives of a choice task, Kaye-Blake et al. (2009) found that more attributes were attended to for the chosen alternative.

Collins et al. (2012) provided a mechanism for showing or hiding the attributes of a large, complex choice task. Up to 22 alternatives were listed, each described by 14 attributes, of which five were always shown, and nine were initially hidden, but could be shown or hidden any number of times prior to choice. The motivation was to provide the respondent with mechanisms for managing the complexity of the choice tasks, and to this end, search[3] and sort tools were also provided. Attributes were shown or hidden at the choice task level. That is, any selection was applied to all alternatives in the choice task. Attribute visibility settings were retained across choice tasks. In their dataset, it could be determined whether an attribute was ever shown, although for practicality, the visibility of the attribute was considered at the time of choice, and parameters associated with hidden attributes were removed from the utility functions for that choice task. Thus, Collins et al. (2012) also handled ANA by observing what information the respondents chose to reveal.

Kaye-Blake et al. (2009) found that in the first choice task, more attribute levels were revealed, and more time was spent on choice. This suggests that the respondent may initially spend more time coming to terms with the available attributes, attribute levels, and tradeoffs, and that this may also be a formative stage with regards to decisions about which attributes to attend to. Supporting this are the findings of Rose and Black (2006), who showed that the degree of heterogeneity related to attributes was a function of choice task completion time, and that most time was spent on the first few choice tasks. A model that only handled the attributes shown to a respondent was compared to another in which all attributes, shown

[3]Alternatives could be eliminated that failed to meet respondent specified criteria.

and not shown, were entered into the utility expressions. The former model outperformed the latter. A comparison of WTP measures was not informative, as no values were significant in either model. Collins et al. (2012) estimated a model that contained responses both from conventional SC choice tasks, and the complex choice tasks that provided the respondent with control over attribute visibility. They found that the latter exhibited greater scale, and so lower error variance. However, the contribution of the attribute visibility tool to this result could not be separated from that of other differences between the two choice task formats, such as the presence of sort and search tools.

These alternative ANA elucidation methods raise a number of issues. One concerns the extent to which ANA is known by the analyst. If an attribute level was never shown, then it is reasonable to assume that the attribute level was ignored[4]. Nonetheless, Kaye-Blake et al. (2009) estimated a model in which the average value of the 'ignored' attribute was entered into the utility expression, and found an improved model fit. This suggests that the respondents might have been making inferences about the likely attribute level, without directing effort to revealing the precise level. Also, even if an attribute was revealed, it is not certain that it was attended to at the time of choice, and integrated into the overall utility of the choice alternative. It may have been subsequently ignored, or it may have been revealed by an errant mouse click. Therefore, information about which attribute levels were selected to be revealed by the respondent cannot be completely relied upon as representing all ANA behaviour. Also, the imposition of a search cost to reveal an attribute level (i.e., clicking on the attribute level in Kaye-Blake et al. (2009) or selecting the attribute from the hide/show tool in Collins et al. (2012)) might be somewhat artificial, and so any findings might confound genuine ANA behaviour with a response to the mechanisms introduced to capture attendance. For example, a respondent may be less likely to attend, due to the costs of clicking on attribute levels. Kaye-Blake et al. (2009) noted that each click may only take a fraction of a second. However, such actions must be repeated many times to reveal all information in a choice task, and this cost may not be viewed as trivial by the respondent. Such temporal costs may have been measured but were not disclosed. Overall, while these methods provide some insights into ANA, they are unlikely to be appropriate as a means of collecting ANA in most studies.

Whereas most studies in the literature have focused on issues such as WTP and model fit, only a few studies have considered the impact of ANA on the scale. Scale has an inverse

[4]This means that one of the models in Kaye-Blake et al. (2009), in which all attribute levels were entered into the utility function, regardless of whether they were revealed, is somewhat questionable.

relationship with error variance, and so differences in scale vary how deterministic the choice model is. Campbell et al. (2008) parameterised the scale, alternately on whether a respondent ignored any attribute, and on the number of attributes ignored. In both cases, scale was normalised to one for respondents that attended to all attributes. They found that scale was larger for those that ignored at least one attribute than for those that did not. When separate scale parameters were estimated based on how many attributes were stated as ignored, the relative magnitude of the scale varied across these parameters, depending on whether the taste coefficients were constrained to zero or not for attribute nonattenders. Nonetheless, scale was always larger for those that ignored some number of attributes, relative to those with full attendance. Accommodating scale also lead to an improved model fit, and further reductions in WTP beyond those observed once the taste coefficients were constrained to zero.

Meyerhoff and Liebe (2009) also estimated scale as a function of whether the respondent stated that they ignored any attributes, or not. They observed attribute nonattenders to exhibit less scale than attenders, when a cross sectional model was estimated. When the panel nature of the data was accounted for, however, no difference in scale was evident. In their study, Meyerhoff and Liebe (2009) handled ANA at the choice task level. They compared their findings to those of Campbell et al. (2008), which utilised serial ANA, and hypothesised that serial ANA might contribute to difference in scale. Overall, whilst it appears that ANA has some interrelation with scale, the findings are not definitive. More research needs to be done, although this line of research will not be pursued in this thesis.

Constraining taste coefficients to zero for respondents that indicated that they did not attend to the associated attribute relies on the assumption that stated ANA responses are completely accurate. However, the literature strongly calls this assumption into question. At a very fundamental level, within the psychology literature, Nisbett and Wilson (1977) present evidence that suggests that when people report on their cognitive processes, they do not do so accurately, but instead use causal theories or judgements based on plausibility.

The most frequent specific evidence from the choice modelling literature comes via the estimation of separate coefficients for those respondents that indicated that they ignored an attribute, and for those that indicated attendance. The coefficients for the stated 'ignorers' overwhelmingly tend to be of correct sign, and lesser magnitude than for the stated attenders (Hess and Rose, 2007; Hess and Hensher, 2010; Campbell and Lorimer, 2009). Some papers have found just some of the sensitivities to the 'ignored' attributes to be significantly different to zero (Campbell and Lorimer, 2009; Alemu et al., 2011), while others have found significance for all attributes (Hess and Rose, 2007; Hess and Hensher, 2010). Where random parameters are employed, taste heterogeneity often remains among stated ignorers (Campbell and Lorimer, 2009; Hess and Hensher, 2010), and Hess and Hensher (2010) found that taste heterogeneity of stated attenders mostly decreased as well. A further advantage to estimating sensitivities for stated ignorers, instead of just constraining the coefficients to zero, is a widely observed improvement in model fit (Hess and Rose, 2007; Campbell and Lorimer, 2009; Hess and Hensher, 2010).

Campbell and Lorimer (2009) additionally found the conditional parameter estimates for stated ignorers to be greater than zero for all attributes except cost, but they neglected to report the variance of the conditional parameter distributions. High variance might weaken this finding. For example, when Hess and Hensher (2010) estimated conditional parameter estimates without consideration for stated ANA, they found that many conditional distributions with a near zero mean had a very high variance.

Carlsson et al. (2010) achieved a similar finding with a single random parameter for each attribute, plus an interaction of that attribute and a dummy variable set to one if the attribute was stated as ignored. The interaction terms were not significant for most attributes, including cost. This suggests that for most attributes, the sensitivities for stated ignorers were not only different to zero, but also not significantly different from those of stated attenders. This approach would be equivalent to the estimation of separate parameters if fixed coefficients are employed, but with random parameters it is subtly different, as the interaction is only entered with the mean of the distribution. The use of separate parameters for stated ignorers and attenders allows a different standard deviation or spread to be estimated for each segment.

Alemu et al. (2011) went beyond estimating just a single sensitivity for stated ignorers. They additionally elicited from the respondents the reasons why they ignored an attribute, and estimated a sensitivity for each, so answering a call for further research on this made by Hensher and Greene (2010). Their findings are an important reminder that attributes may not be attended to for a multitude of reasons[5]. When an attribute was ignored because the respondent did not find the attribute important, the coefficients were never significantly different to zero. When a respondent ignored an attribute because they consequently found it easier to choose between alternatives, significant sensitivities were estimated for five out of six attributes. This dropped to three out of six attributes both when the respondent found the attribute levels unrealistically high or low, and when they believed that the attribute

[5]Recall the discussion in Section 2.2.

should not be traded against other attributes. The sensitivities were significant for four out of six attributes that respondents stated they ignored for reasons unknown to them.

The evidence from Alemu et al. (2011) suggests that stated ignorers may not *just* put less weight on the attribute (Carlsson et al., 2010), or ignore the attribute for only some choice tasks (Hess and Hensher, 2010). Instead, some respondents may *consistently and truly* ignore the attribute, most likely if they were motivated to do so due to true indifference to the attribute. Thus, relying on a binary indicator of nonattendance, and estimating coefficients for each of the two attendance conditions, is a form of aggregation bias, and may mask true ANA. In a sense, this is a reversal of the problem whereby not handling ANA leads to a failure to recover the significant sensitivity to the attribute by a subset of individuals, as was experienced for example in Rose et al. (2005). In that case, true sensitivity was masked, while in the case at hand here, the estimation of sensitivities to stated ignorers might mask true *in*sensitivity.

Further evidence to suggest that stated ignorers may not all just have a moderated sensitivity to the attribute can be observed in Campbell and Lorimer (2009) and Hess and Hensher (2010). They estimated random parameters for stated ignorers, and for a number of attributes found the standard deviation of the RP distribution to be relatively large, compared to the mean. For example, in Hess and Hensher (2010), the random parameter for stated ignorers of free flow time had a mean of -0.1357 and a standard deviation of 0.1330. This implies that 15.38 percent of coefficients were positive, and of implausible sign. While the normal distribution will always have unbounded support, it may be that a mass of true ignorers led to such a high percentage of implausibly signed coefficients. Hensher (2007) found that ANA might lead to sign violations, and this finding will be confirmed and explored in this body of work, in Chapter 4, and mitigated, in Chapters 5 and 6. Thus, the distribution might be attempting to capture a point mass at zero representing true ANA, plus some range of true sensitivities.

Balcombe et al. (2011) have provided a counterpoint to the many studies that have suggested that stated ANA might suffer from reliability problems, and that freely estimating coefficients for those individuals that state that they ignore an attribute[6] leads to improved model fit. They compared the two approaches in a model estimated in the Bayesian framework, using the marginal likelihood as a way of comparing non-nested models. It was found that constraining the coefficients to zero led to a better model fit than via interactions, in three out of four fundamentally similar variants of their choice experiment.

[6]Either through two separate main effects parameters, or via an interaction, as discussed.

Hensher et al. (2007) also recognised the limits of studies that deterministically handle stated ANA information. They argued that it is more realistic to assume that "the exogenous information points to the correct likelihood specification for a respondent with error" (p.75). That it, stated ANA cannot be relied upon to be completely accurate. They proposed an estimation procedure that allows stated ANA to be handled stochastically rather than deterministically. First, a choice model was estimated, wherein the choices were the combinations of stated nonattendance across the attributes, as elicited from the respondent. The utility expressions were specified as a function of age, income, and the attribute levels of the choice tasks. The expected maximum utility (EMU) was calculated for each respondent, and sequentially introduced into a second model, where the choice alternatives were the alternatives of the choice task. That is, the first model handled the decision making process, and the second model handled the choice conditional on the process employed. Significant interactions were found between the EMU and the mean of two of the attributes. Model fit, as measured by the ρ^2, improved. The VTTS increased once ANA was accounted for, where the difference in both the mean and variance of the measure was found to be significant. Whilst this approach does not assume that stated ANA is completely accurate, it is still reliant on stated ANA data.

There has been some speculation as to why stated ANA may be inaccurate. As noted previously, Alemu et al. (2011) found that the degree of accuracy varied across a number of reasons that the respondent could nominate: genuine indifference to the attribute, as a way of reducing cognitive burden, implausible attribute levels in the choice tasks, as a protest against having to trade off certain attributes, and 'don't know', which might have been treated by the respondent as 'some other reason'. Only the first of these reasons led to marginal utilities that truly were zero. Thus, all other motivations may compromise the accuracy of stated ANA. Some have suggested that the attribute may only be ignored in some choice tasks (Carlsson et al., 2010). This may be for a variety of reasons, including as a consequence of the combination of levels presented (Hess and Hensher, 2010), to break ties (Balcombe et al., 2011), or because different decision making strategies, beyond just attribute ignoring, may be employed in different contexts and at different stages of the decision process (Hensher et al., 2007; Stewart et al., 2003). Alternatively, there may be social pressures, and the respondent may believe that a 'good respondent' would attend (Balcombe et al., 2011). Irrespective of the reason, the evidence of unreliability serves as a warning against assuming otherwise.

A further criticism of stated ANA has been levelled on endogeneity grounds. Hess and Rose (2007, p.23) noted that "the information on IPSs cannot be regarded as an independent variable", where they elicited from respondents information on two such information processing strategies (IPSs): ANA, and the aggregation of common-metric attributes. Consequently, the stated ANA variables may be correlated with the unobserved factors, and an endogeneity bias may be engendered. No test was provided, however. To overcome this issue and yet still exploit stated ANA, Hensher (2008) proposed a joint model of process and outcome, wherein it is recognised that choices are made conditional on the processing rules adopted by the respondent. Each alternative modelled was the joint choice of an attribute processing rule, and the choice of alternative conditional on this processing rule. The attribute processing rule was defined as a unique combination of attendance and nonattendance to the attributes describing the choice alternatives. Separate taste sensitivities were estimated for each attribute processing rule. While this approach represents an advance, it is still dependent on stated ANA, which may be unreliable, and which requires that such information be obtained, where this may be more costly, or not possible for existing datasets. Additionally, the approach is unlikely to scale well as the number of attributes increases, due to a proliferation of parameters.

Overall, a number of concerns have been raised that call into question the viability of using supplementary questions concerning ANA behaviour as a way of accommodating ANA in choice models. There are concerns about endogeneity, and with regards to the reliability of the responses, Hensher (2008, p.299) concedes "we will never be able, with total certainty, to rely on a set of exogenous data items to elicit how an attribute is processed by each individual". The literature has responded by developing a number of analytical methods, whereby an appropriately specified model can infer and accommodate latent ANA behaviour.

Latent class methods

Latent class models have been used to identify ANA in two ways. One way is through the interpretation of insignificant taste coefficients in a class as ANA (Swait and Adamowicz, 2001). The other way is through the censoring of taste coefficients to zero in certain classes (Hess and Rose, 2007).

Swait and Adamowicz (2001) implemented a variant of the conventional LC model, wherein class assignment was driven by an entropy measure, which served as a summary measure of choice task complexity. Two classes were estimated. One class contained taste coefficients for product attributes that were mostly significant, and so represented close to full information processing. The other class contained many insignificant taste coefficients for product attributes, and a greater difference in brand constants. This class was interpreted as a brand based decision strategy, in which many of the product attributes were not considered. This technique has the appeal of associating ANA with a certain type of choice behaviour. However, the use of conventional LC models to recover ANA may not be effective if ANA for one attribute varies independently of the tastes and ANA of the other attributes.

One analytical method that has gained traction in the literature as a way of inferring ANA is a variant of the LC model, referred to herein as the LC approach[7]. Hess and Rose (2007) were the first to implement such models. They estimated a series of LC models, each of which tested for nonattendance to one of the attributes in the choice tasks. Two classes were specified, and crucially, in one class, the taste coefficient for one of the attributes was constrained to zero, and so not estimated[8]. This class represents nonattendance to the attribute, and the class assignment probability can be interpreted as the ANA rate for the attribute in question. A single constant can be introduced into the class assignment component of the model, resulting in the same ANA rate for all individuals. Alternatively, covariates can be introduced into the class assignment component of the LC model, thus allowing the ANA rate to vary across segments of individuals. Hess and Rose (2007) employed the second approach, and found significant differences in ANA between individuals based on their age and income, as well as some changes in VTTS values from the baseline MNL model. Hess and Rose (2007) and Hensher and Greene (2010) also used the LC approach to differentiate between respondents who added up common-metric attributes such as the toll and running costs associated with a car trip (Hensher, 2004; Layton and Hensher, 2010). This demonstrates the versatility of the LC approach for capturing a range of choice heuristics or processing rules. However, if nonattendance is to be inferred for more than one attribute, or multiple heuristics are to be handled, then the parametric complexity of the model will increase, and the analyst must make a number of important decisions regarding model specification.

One such decision is the number of classes to specify. If nonattendance is to be modelled for $K^\star$ attributes, and no other heuristic is to be modelled, then up to $2^{K^\star}$ classes can be specified. Hensher et al. (2012a) noted that each class has a specific meaning, unlike with a conventional LC model, where each class represents some *arbitrary* combination of sensitivities across the attributes. Instead, each class represents some *specific* combination of

[7]The approach will be described in general terms here. It forms the basis the for the model proposed in this thesis, and so will be formally notated in Chapter 3.

[8]Scarpa et al. (2009) recognised that if an attribute uses dummy or effects coding, then *all* coefficients associated with an attribute must be constrained to zero.

ANA across the attributes. Both behavioural and econometric problems may result from the exponential relationship between the number of attributes for which ANA is to be modelled and the number of classes. Behaviourally, only some of the potentially very large number of ANA combinations may actually be employed by decision makers. This may also lead to an econometric problem, as when a conventional MNL model is used to control class assignment, the number of class assignment parameters will increase exponentially. This may lead to estimation problems, especially for combinations of ANA that are employed by individuals with low frequency. Indeed, only two studies have retained all possible ANA combinations (Campbell et al., 2010b; Hensher et al., 2012a). Campbell et al. (2010b) examined a choice context with five attributes, resulting in 32 classes. Notably, some combinations had extremely low probability. Hensher et al. (2012a) handled ANA for three attributes, with eight classes. All other studies that have used the conventional LC approach have retained only a subset of the ANA combinations (Scarpa et al., 2009; Hensher and Greene, 2010; Campbell et al., 2011; Scarpa et al., 2011).

In deciding how many and which classes to specify, Scarpa et al. (2009) tested to see if adding more classes would improve model fit. Conversely, Scarpa et al. (2011) employed a 'general-to-specific' specification search, starting with all classes, and progressively dropping those that had a probability of less than three percent. Unsurprisingly, the decision of which classes to retain appears to be crucial to the performance of the model. Scarpa et al. (2009) found that some specifications led not only to worse model fit than for their final accepted model, but also wildly divergent model outputs, such as ANA rates for each attribute. Hensher et al. (2012a) found that eliminating some ANA patterns (i.e., classes) led to counterintuitive model outputs, such as implausible relative magnitudes of estimated taste coefficients.

A further decision that must be made by the analyst is whether to constrain the coefficient(s) for each attribute to be the same across all classes (when not constrained to zero), or whether to estimate separate coefficients for each attribute in each class. The first approach is sometimes referred to as the equality constrained latent class (ECLC) approach (Scarpa et al., 2009), in partial recognition that when not constrained to zero, the same taste coefficient can be specified for an attribute by constraining it to be equal across classes. This approach has also been employed by Campbell et al. (2010b), Scarpa et al. (2011) and Hensher et al. (2012a). One advantage of this approach is that it is more parsimonious than the alternative, as the *taste* coefficients do not proliferate as more classes are added, and the parametric cost of more classes comes only through the class assignment component of the model. Campbell et al. (2010b) noted that only process heterogeneity is captured, while Scarpa et al. (2009) acknowledged that estimation of separate coefficients for each class allows taste heterogeneity to be captured across the classes, and that this may be one of the analyst's objectives. This second approach was taken by Hess and Rose (2007), Hensher and Greene (2010) and Campbell et al. (2011). It is parametrically more expensive, as if there are $K^\star$ attributes for which attendance is to be modelled, then each additional class entails the estimation of up to $K^\star$ more taste coefficients. Specifically, if all combinations of ANA are handled, $\frac{2^{K^\star}K^\star}{2}$ taste coefficients must be estimated. Given that handling all combinations of ANA will require $2^{K^\star}$ classes (more on this below), the total number of parameters to estimate quickly becomes prohibitive.

Also, when the coefficients can vary across classes, there is a potential confoundment between the ANA behaviour, which the analyst seeks to reveal, and taste heterogeneity. Ideally, unique coefficients in each class would reflect the preferences of respondents that adopt that pattern of ANA behaviour that the class implies. However, this assumes that the pattern of zero constraints across the attributes in the class has a dominant influence on the estimated likelihood. It is possible though that despite the zero constraints, the class primarily captures some pattern of taste heterogeneity across the attributes, as in a conventional LC model, and that this taste heterogeneity has a more dominant influence on the likelihood than the process heterogeneity that is reflected in the zero constraints. Further, it is plausible that this confounding may be more pronounced for classes that represent nonattendance to fewer attributes (i.e., have fewer zero constraints), as there are more coefficients freely estimable, and so there is greater scope not just to capture taste heterogeneity, but also correlation in tastes across these attributes. As a counterpoint, constraining the coefficients across classes when the true sensitivities vary across different expressions of ANA may increase the error associated with the class assignment probabilities. On balance of these two arguments, this body of work constrains the taste coefficients across classes, but allowing them to vary remains an option, albeit a parametrically expensive one. The potential confoundment between ANA and taste heterogeneity does not consequently recede, as will soon be made clear in this literature review. However, this thesis will make a contribution by demonstrating through simulation the extent of this confounding, and by allowing the taste heterogeneity to be captured elsewhere in the model.

An alternative LC approach has been proposed by Hole (2011a), who called it an 'endogenous attribute attendance' (EAA) model. The approach seeks to overcome the problem of parameter explosion encountered by the conventional LC approach, whereby each class, bar one, requires at least one parameter to be estimated, and handling all combinations of ANA may become impractical. Within the EAA model, a binary logit model for each attribute for which ANA is to be modelled controls the probability of whether that attribute is attended to or not. In its simplest form, each binary logit model contains a constant only, and thus the class assignment component of the model requires only as many parameters as there are attributes for which ANA is to be modelled, $K^{\star}$. The approach is therefore more parsimonious than the conventional LC approach, and can more readily handle all combinations of ANA. The parsimony does not preclude a more complex model specification, as covariates can be introduced into the binary logit models to vary the probability of ANA across respondents. If ANA to only one attribute is handled, the approach is analogous to that of Hess and Rose (2007). However, for larger values of $K^{\star}$, the probability of each combination of ANA is the product of the appropriate probabilities (attendance or nonattendance) for each of the $K^{\star}$ attributes. Since this model will be extended upon in Chapter 3, a detailed, fully notated explanation of the model will be provided there.

The EAA model relies on the assumption that ANA is independent across the attributes. Therefore, it may struggle to handle situations in which specific combinations of ANA are employed with disproportionate frequency by respondents. Hole (2011b, p.4) noted that in the conventional LC approach, the choice by the analyst of which ANA subsets to consider is not 'obvious *a priori*', and so concluded that it is an advantage of the EAA model not to have to make such a choice. Whilst the appeal is intuitive, the independence assumption upon which the model rests may not hold, resulting in model misspecification, and potentially erroneous inferences being made. Ease of specification of a model is indeed an advantage, but does not absolve the analyst from taking adequate measures to ensure the model's appropriateness.

Hole (2011a) obtained empirical results similar to much of the literature that has employed the conventional LC approach, and so these will be discussed together below. Notably though, the EAA model outperformed both the MNL and RPL models (Hole, 2011b), but not the non-parametric, discrete density model of Bajari et al. (2007) and Train (2008), wherein a very large number of coefficients are fixed, and only the shares for these coefficients are estimated. This is likely due to the extreme flexibility of the latter approach, which hints at the potential gain in specifying a very flexible model that can also handle ANA. The LC approach has limited flexibility, since only a single point mass is estimated to reflect the sensitivity to each attribute, conditional on attendance.

Hole et al. (2012) extended the EAA model to include an additional class of response behaviour, for those that attend to all attributes. This introduced some degree of taste heterogeneity, and allowed for an elevated incidence rate for full attendance, thereby partially overcoming one of the drawbacks of the assumption of independence of ANA. Stated ANA responses were also introduced as covariates in the binary logit models controlling the inferred ANA rate, thus allowing stated ANA to be handled probabilistically. Hole (2011a, p.204) noted that an extension to include random parameters would be "conceptually straightforward (but computationally intensive)". Such an extension forms one of the contributions of this thesis. It will be shown that there are some important issues to be resolved before such a model can be reliably estimated.

All studies that have implemented the LC approach to ANA have found an improvement in model fit over an MNL model. Hensher and Greene (2010) and Hole (2011b) found that it also outperformed an RPL model. Scarpa et al. (2009) observed the parameter estimates to be more significant than in an MNL model.

Many studies have found ANA rates to be high, sometimes disturbingly so. Cost has been found to be particularly susceptible, and particularly disturbing, due to the key role that the attribute plays in the formulation of WTP values. Nonattendance to cost precludes the generation of WTP for those that do not attend (Rose et al., 2005; Carlsson et al., 2010; Hensher et al., 2012a). Campbell et al. (2011) found that over 60 percent of respondents did not attend to cost, while the figure was 70.5 percent for Hole (2011a) and over 75 percent for Campbell et al. (2010b). The ANA rate for cost was higher again for Scarpa et al. (2009), at 90.9 percent. They also inferred a similar, if slightly muted ANA rate from a stochastic attribute selection model (introduced shortly, in Section 2.2.2), at 83.2 percent. In contrast, Hensher and Greene (2010) found ANA rates to be generally lower, and notably so for cost attributes, with an ANA rate of just 3.6 percent for a road toll, and 4.7 percent for automobile running costs. The difference might stem from the type of choice task, with various driving costs being broadly experienced, and not contentious, albeit disliked. Three of the above were non-market valuation studies (Scarpa et al., 2009; Carlsson et al., 2010; Campbell et al., 2010b), in which the respondent might have trouble trading off attributes with cost, especially for "unfamiliar ecological and environmental goods" (Campbell and Lorimer, 2009, p.4). It might be that the respondent does not believe that they will directly have to pay for the change from the status quo.

As with studies employing stated ANA, there does not appear to be a consistent direction of change in WTP values once ANA has been accounted for through the LC approach. Some have observed an increase in WTP (Hensher and Greene, 2010; Hensher et al., 2012a), others an insignificant decrease (Scarpa et al., 2011), others a mild decrease (Hole, 2011a; Hole et al., 2012), and others still a large decrease (Scarpa et al., 2009; Campbell et al., 2010b, 2011). The large decreases resulted in much more plausible WTP values, and appear to be the result of very large rates of nonattendance to cost. Handling this nonattendance to cost greatly increased the marginal disutility of cost for those that did attend, thus increasing the magnitude of the WTP denominator and decreasing WTP.

Scarpa et al. (2011) exploited the ability to condition the class assignment probabilities of the LC model on the observed sequence of choices of each respondent (refer to Section 2.1.2 for details). They wished to infer which specific respondents were not attending to an attribute, and then investigate what was systematically influencing the nonattendance. The individual specific class assignment probabilities[9] were summed for all classes which represented each attribute as being ignored, to generate an overall individual specific probability of ANA. If this probability exceeded 50 percent, the individual was classified as not attending to the attribute. The problem with this approach is that 50 percent is not a very rigorous threshold. Indeed, if the class assignment probabilities are used to probability weight the class specific coefficients, then in this study a threshold of 50 percent would represent a coefficient half the magnitude of the freely estimated coefficient[10], which is likely to represent at least a moderate sensitivity to the attribute. A number of factors were found to be significant drivers of this constructed ANA indicator, including gender and the rank of choice in the rank exploded dataset which the paper utilised. However, that a stated ANA response was not significant supports the argument that the indicator is likely capturing more than ANA[11]. Caution is warranted when utilising conditional class assignment probabilities to infer individual specific ANA. At the very least, a larger threshold probability is advised.

Campbell et al. (2012), concerned by the high rates of ANA observed in past studies for some attributes, and for cost in particular, showed that such high ANA rates may be a consequence of not adequately handling taste heterogeneity, and of nonattendance not to the attribute as a whole, but to specific attribute levels. They postulated that low sensitivity to an attribute may be confounded with nonattendance to that attribute. Focus was placed exclusively on the cost attribute, because of the crucial role it plays in deriving WTP measures. Rather than handle preference heterogeneity with a RP distribution, they retained the LC approach, but estimated additional point masses representing sensitivity (i.e., attendance) to

[9] More precisely, the choice sequence specific class assignment probabilities.

[10] The model applied an equality constraint across classes, and so the coefficients for each attribute either assume zero, or the single estimated coefficient.

[11] Stated ANA may simply have been unreliable. However, even a low level of significance may have been expected.

the attribute, in addition to the point mass constrained to zero for which only a share was estimated. The ANA rate for cost reduced from 71.8 percent when a single nonzero point mass was employed, to 64.3 percent when three nonzero point masses we employed. The three estimated coefficients represented mild, moderate, and strong sensitivities to cost. Model fit improved noticeably. Campbell et al. (2012) concluded that there may be confounding between taste and ANA heterogeneity, and that models that can handle both are "better equipped to disentangle respondents who are relatively cost-insensitive from those who did not attend to price" (p.11).

Campbell et al. (2012) also found evidence for the existence of thresholds and cutoffs as respondents evaluated cost. Rather than setting the coefficient to zero for all levels of an attribute, the censoring was imposed for combinations of attribute levels. Once again, an LC approach was employed, with each class representing a combination of censorings across the cost attribute levels, with the coefficient constrained to equality across classes. Fifteen combinations were tested, with up to three nonzero taste coefficients. As more nonzero taste coefficients were successively added, the ANA rate *for all attribute levels* decreased from 59.4 percent (one taste coefficient), to 35.7 percent (two), to just 9.7 percent (three). Therefore, the decrease in nonattendance to cost, as preference heterogeneity is accommodated, is much more pronounced once nonattendance is handled with respect not just to the attribute, but the attribute levels. Unfortunately, the approach employed by Campbell et al. (2012) does not scale well. For cost alone, with four attribute levels and three point masses capturing preference heterogeneity, 48 parameters were required. In conclusion, Campbell et al. (2012) made a broad call for studies incorporating ANA to handle both taste and process heterogeneity; a call that is answered in this thesis.

Hess et al. (2011) extended the LC approach such that conditional on attendance, the sensitivities are estimated with random parameters. This allows taste heterogeneity to be captured through the random parameters, and should lessen the chance that the zero coefficient is approximating low sensitivities. They observed an improved model fit over the RPL model and the conventional LC approach, as well as lower ANA rates than with the LC approach. This thesis proposes a similar, but more flexible model, and provides many findings that extend upon Hess et al. (2011). However, the research in this thesis was conducted independently and concurrently, and the model was developed, implemented and tested before the Hess et al. (2011) working paper was made available. More details about the differences are provided in Section 2.3, which outlines the contribution of this thesis to the literature.

Conditional parameter estimates

Another approach to identifying and accommodating ANA utilises the parameter distributions that are conditioned on the observed choices, in the context of a RPL model. A discussion of the derivation and interpretation of these conditional parameter distributions was provided in Section 2.1.3. Their first application to ANA was by Hess and Rose (2007). They compared the means of the conditional distributions[12] of those individuals that stated they ignored an attribute, and those that stated they attended to the attribute. After deriving a distribution of means across all individuals in the sample, they found that the distribution for ignoring respondents was closer to, but not degenerate at, zero. Further, with one exception, this difference was only small. The authors concluded that while some individuals probably did ignore each of the attributes, the stated ANA responses may not have been accurate.

Hess and Hensher (2010) developed the approach further. They considered more than the conditional mean, by calculating the coefficient of variations of the conditional parameter distributions, which is the mean divided by the standard deviation. It was found that high coefficients of variation were estimated only when the conditional mean was near zero. This implies that these individuals may be ignoring the attribute. Hess and Hensher (2010) suggested a threshold value of two, whereby if the coefficient of variation exceeds this value, the attribute is considered as ignored, for that individual. They noted that the value is somewhat arbitrary, and called for more research into the determination of the value. The final step is to re-estimate the model, with two parameters estimated for every one estimated initially. One parameter is estimated for those respondents for whom it was inferred that they ignored the associated attribute, while the second parameter is for the remaining respondents, who are assumed to all attend to the attribute.

Hess and Hensher (2010) applied the conditional parameter estimate approach to a route choice SC study, in which stated ANA was obtained, allowing comparisons to be made between stated and inferred ANA. Some variation was observed between the two sets of rates. The inferred ANA rate was higher than the stated rate for one attribute, very similar for one more, and lower for three attributes. The alignment at the individual level was not so close, however. For all attributes, more than 50 percent of individuals who stated that they ignored the attribute were inferred as not ignoring the attribute. More encouragingly, the model was found to outperform both the base RPL model, and a RPL model in which separate

[12] A common misunderstanding of the conditional parameter distributions in the RPL model is that they are a point estimate, where in fact they are a distribution. The mean, however, will represent the most likely value for an individual that makes the observed sequence of choices.

parameters were estimated based on *stated*, not inferred, ANA. Also, all parameter estimates representing ignored attributes were insignificant, in contrast to the second RPL model just mentioned, where all separate parameters for stated nonattenders were found to be significant. Finally, individuals for whom attendance was inferred in this model demonstrated less preference heterogeneity than the full sample of individuals in the base RPL model. Hess and Hensher (2010) suggested that preference heterogeneity may in part be an 'artefact' of ANA. This is an important motivation to adequately handle ANA, as it suggests that a continuous distribution of preferences alone may not be sufficient.

The conditional parameter estimate approach shows considerable promise, however, some caveats must be noted. The approach employs a two stage, sequential estimation. First, the nonattenders are identified, and then the final model is estimated. This effectively doubles the computation time, and requires that the dataset be modified between the stages, although this could be automated with software. Typically, a final RPL model is arrived at after a nontrivial specification search, including the testing of a variety of RP distributions. The most appropriate comparison of specifications would be performed after nonattendance has been handled. Therefore, the two stages would need to be completed for every model specification, adding to what can already be a heavy burden on the analyst of a multitude of model specifications, each with long estimation times.

Also, since the accuracy of the conditional parameter distributions is dependent on the length of the panel (Train, 2009), the accuracy with which the ANA rates are identified are likely to also depend on the panel length. Indeed, this was found to be the case by Mariel et al. (2011), who used simulations to reveal a deterioration in accuracy of inferred ANA as the panel length decreased. However, obtaining more than a small number of observations per individual may be costly[13], and in some situations, especially with revealed preference data, may be impossible. Additionally, if ANA behaviour is not consistent for an individual over choice tasks, then the technique would be likely to reveal some sensitivity to the attribute, even if the attribute was ignored in some choice tasks.

Hess and Hensher (2010) recognised that the choice of threshold is somewhat arbitrary. Mariel et al. (2011) found, again using simulated data, that the most accurate threshold differs as the true ANA rate differs. This is obviously problematic, since the ANA rate is the very thing that the analyst is attempting to estimate. The analyst might have a rough sense of what the true rate will be, but it is nonetheless latent. More detail on the simulations

[13]The number will vary based on the complexity of the choice task, and how long each choice task takes to complete.

of Mariel et al. (2011) is provided in Section 2.2.3, where the use of simulated data in the context of ANA is reviewed.

Mariel et al. (2011) also reported an empirical application of the conditional parameter estimate method, to the valuation of landscape externalities of wind power generation. The most notable feature is the way that they handled 'controversial attributes' - those attributes, such as the height of wind turbines, which some people have a preference for, and others have a preference against. They noted that the mass of coefficients near zero, which are a consequence of a RP distribution spanning zero, precludes the identification of nonattenders using the conditional parameter estimate method. Their solution was to first determine if each respondent has a positive or negative coefficient, and then estimate separate parameters for each, prior to applying the coefficient of variation threshold.

Censored random parameter distributions

A methodology that has not received sufficient attention in the ANA literature for its ability to capture ANA is the censored normal RP distribution, as proposed by Train and Sonnier (2005). They discussed a range of contexts in which a constraint might need to be imposed on the sign of a RP distribution. However, they also noted that such a bound may be required for a "desirable attribute that is valued (or, at worst, ignored) by all customers" (p.2). That is, they explicitly noted the behavioural validity of an individual ignoring an attribute.

Train and Sonnier (2005) proposed two transformations of the normal distribution, beyond the widely adopted lognormal distribution. One was Johnson's S_B distribution, and the other the censored normal distribution. The latter draws from a normal distribution, and either retains the draw if it is of plausible sign[14], or censors the value to zero, resulting in a point mass at this value. They noted that the censored normal distribution is useful for "an attribute that some customers do not care about (i.e., are indifferent to its presence and simply ignore) and other customers find desirable" (p.4). Clearly, the authors were motivated to handle ANA. As they introduced the technique, they appeared to be focused on ANA as a form of preference heterogeneity. Nonetheless, when they observed ANA rates such as 51 percent for an increase from low to mid-level performance, in an automobile choice context, they speculated that it might be a consequence of the levels included in the choice experiments. That is, it might in part be an artefact of the experimental design. Train and Sonnier (2005) observed high levels of ANA, and a considerable improvement in model fit, when they introduced the censored

[14] As defined by the analyst, either positive or negative depending on the context.

normal distribution in their empirical application.

The use of censored normal distributions in a RPL model is conceptually simple, readily applied, and easily interpreted. Unlike the stated ANA methods outlined in the previous section, it is an analytical method that infers ANA, and so has no additional data requirements. The key limitation of the method is that it risks confounding ANA with preference heterogeneity. To see why, consider a censored normal distribution, wherein the underlying normal distribution has a mean of μ and standard deviation of σ, with the second moment clearly controlling the extent of preference heterogeneity. However, the share at zero is $\Phi(-\mu/\sigma^2)$, where Φ is the standard normal cumulative distribution. Therefore σ is impacting upon both the continuous distribution of preference heterogeneity, and the size of the point mass at zero. Empirically, this may well approximate the true continuous distribution of preference heterogeneity, as well as the ANA rate. However, the confounding may also limit the ability to accurately estimate each. Essentially, two parameters are being employed to recover three aspects of the distributions: the mean, the standard deviation, and the size of the point mass at zero. This thesis will overcome such a restriction, albeit at a computation cost.

Mixtures of distributions

Fosgerau and Hess (2008) compared four widely used RP distributions with two semi-parametric alternatives: the Legendre polynomial approach of Fosgerau and Bierlaire (2007), and a mixture of distributions approach (Revelt, 1999). These alternatives provide more flexibility in a RP distribution by allowing an arbitrary number of structural parameters to be specified, limited of course by the size of the dataset. Notably, a mixture of distributions can accommodate ANA. Under this approach, for each attribute, multiple continuous distributions are estimated, as well as the discrete probability of each of these distributions. The number of distributions is specified by the analyst. Fosgerau and Hess (2008) utilised normal distributions, but others could be employed. If the standard deviation of one of the distributions tends to zero, that distribution becomes degenerate, and represents a point mass. Fosgerau and Hess (2008) noted that most point masses are not plausible, except for zero, which can represent the ignoring of or indifference to an attribute.

Fosgerau and Hess (2008) first used simulated datasets to compare the performance of the distributions. One motivation was to determine if a point mass at zero, representing ANA, could be recovered using a mixture of normals. The true distribution was specified as a normal, with probability of 80 percent, and as a point mass at zero, with probability of 20 percent. The results were not definitive. One of the distributions did not become degenerate, but a steepening of the cumulative distribution function near zero hinted at the presence of the point mass. Of concern was a higher variance across simulation replications than the other distributions.

Mixtures of distributions were also employed on an empirical dataset by Fosgerau and Hess (2008), in the context of a VTTS study. Such distributions proved to be one of the best in terms of model fit. Degeneracy was observed for three out of four attributes, each close to, but not at zero. The authors noted that the degeneracy appeared to occur "where the true density places a lot of mass, even if it is unlikely to be point masses" (p.14). Campbell et al. (2010a) also employed a mixture of normals for a single attribute, in an empirical setting. Model fit improved, but although one of the two distributions was close to zero, it did not become degenerate. Whilst flexible distributions such as mixtures of normals display some promise, including the *potential* to identify ANA, this potential does not appear to be realised when tested using both simulated and empirical data.

Stochastic attribute selection models

Motivated by such possible causal factors as cognitive constraints, simplifying heuristics, and above all, the consumer that values only some attributes of a product that can be described by many, Gilbride et al. (2006) developed a technique for handling ANA that is grounded in the Bayesian variable selection procedure of George and McCulloch (1993). The earlier work was concerned with aggregate linear models, whereas Gilbride et al. (2006) extended this approach to handle the selection of attributes at the individual level, in a Bayesian random parameters discrete choice model. That is, individuals may select just some of all available attributes when making choices; all other attributes are ignored. The reader is referred to Scarpa et al. (2009) for a concise explanation of the mechanics of the model, and to Gilbride et al. (2006) for a full exposition.

As with other Bayesian models, posterior distributions are generated for each parameter for each individual. However, the difference here is that the distribution can also contain a point mass at zero, which cannot be achieved with conventional RP distributions. Gilbride et al. (2006) suggested that the mode of the individual specific posterior distribution can be used to classify whether that individual attended to an attribute or not. If the mode is the point mass at zero, the individual ignored the attribute. In other applications, they recommended integrating over the entire distribution, to handle uncertainty in the inference of ANA. Gilbride et al. (2006) retained the terminology of variable selection. In their application, Scarpa et al. (2009) instead referred to it as stochastic attribute selection. The latter terminology will be employed here, as it more clearly communicates that it is *attributes* that are or are not considered by an individual.

Gilbride et al. (2006) tested the stochastic attribute selection model on simulated data. They were able to recover the true parameters accurately, although with less accuracy for true values that were closer to zero, suggesting that there may be a small degree of confounding between the point mass and the continuous distribution. They also tested the model empirically in a marketing context, on products that could not be disclosed. Nineteen choice tasks were presented to each respondent, with ten described by a full 16 attributes, and nine described by only nine attributes. These were referred to as full and partial-profiles, respectively. A variety of model specifications were tested to see if the number of attributes had an impact on the incidence of ANA. Best model fit was achieved by a model in which ANA was only modelled for the full-profiles. The best predictive power was achieved with a model where ANA was modelled for both types of profiles, but where any attribute not attended to in the partial-profiles could not be attended to in the full attributes. They found that the ANA strategy imposed by respondents varied across choice contexts, as the level of complexity varied. Of the 16 attributes in the full-profiles, 45.5 percent were ignored on average. The parameter estimates were of greater magnitude than the baseline model, suggesting that ANA biased the estimates downwards, which is plausible when the influence of ANA is not separated out. Contrary to other findings such as Hess and Hensher (2010), the preference heterogeneity increased from the baseline model. It was also found that for those that attended to an attribute, a simulated change in product configuration led to a strong impact on choice probability; much more so than if full attendance was assumed for all.

Scarpa et al. (2009) also estimated a stochastic attribute selection model, in the non-market context of rural landscape valuation. They observed a range of findings consistent with Gilbride et al. (2006): improved model fit over the baseline Bayesian model, a decrease in magnitude of the means of the taste coefficients, and an increase in the standard deviation. As with an LC model that they also estimated, they obtained drastically reduced WTP values once ANA was accounted for. Some discrepancies were observed in the ANA rates obtained from the two types of models, but they were roughly aligned, and it is not possible to determine which was more accurate.

The stochastic attribute selection model shows some promise. It allows ANA to be determined analytically, and so is not reliant on stated ANA responses. Also, no assumptions need to be made about the structure of ANA across attributes. For example, the frequency with which certain combinations of ANA are observed can be determined directly from the posterior distributions (Scarpa et al., 2009), and the analyst does not need to make assumptions about what combinations are valid. However, estimation of the model is limited to the Bayesian framework. This impedes its applicability in practice, due to the widespread use of classical estimation. Further, Bayesian estimation does not handle certain model specifications well, such as those that include fixed coefficients, and distributions with bounded support, such as the triangular distribution[15] (Train, 2009).

Behavioural theoretical models and other approaches

Several other approaches for capturing ANA have been proposed. DeShazo and Fermo (2004) considered a range of variables which they believed systematically influenced ANA. Two-way interactions were generated between these variables and the attributes for which it was believed they influenced nonattendance. Then, in addition to estimating a main effect for each attribute, each interaction was parameterised. If an interaction parameter was found to be significantly different from zero it was concluded that the variable influences the propensity to attend.

Cameron and DeShazo (2011) handled these systematic influences in a more flexible way. For each attribute, they generated a single propensity to attend measure, which was multiplied by the estimated taste coefficient. This measure was a transformation of an estimated multivariate index, thus allowing multiple influences on the propensity to attend to be captured. Several transformations were proposed. Adding the index to one is the simplest, but does not guarantee preservation of sign of the associated taste coefficient. An alternative transformation that does preserve the sign is exponentiation of the index, which forces the propensity to attend measure to be positive. Finally, the logistic cumulative density function constrains the measure to be between zero and one, which is even more appealing.

Hensher and Rose (2009) also multiplied a propensity to attend measure by the taste coefficients. However, rather than rely upon systematic influences, they integrated this measure over a density, thus allowing a random distribution of propensities to attend across individuals. This approach drew upon a similar methodology employed by Layton and Hensher (2010) in the context of common-metric attribute aggregation.

All three of the above studies assumed preference homogeneity. Of concern is whether

[15] Bayesian estimation also has some advantages over classical estimation, including easy estimation of correlated normals, which leads to a proliferation of parameters under classical estimation.

the systematic and randomly varying propensities to attend are indeed capturing ANA, or just preference heterogeneity. For example, Cameron and DeShazo (2011) found that the systematic sources of propensity to attend that were significant when preferences were assumed to be homogenous were no longer so once preference heterogeneity was introduced into their model.

Cameron and DeShazo (2011) is notable in that they specified a behavioural theoretical model, in contrast to most other studies, which largely treat ANA statistically. Their model attempts to identify specific influences on ANA, and suggests that the decision to attend to an attribute is influenced both by the benefits and costs of attention. Notably, the benefits from attending to an attribute are a function of the expected utility loss resulting from a suboptimal decision that is made if the attribute is ignored. The model implies that the ANA rate is a function of the composition of the choice set, and hence, in the context of stated choice experiments, the experimental design.

Arentze et al. (2011) proposed a theoretical model that allows the attendance to an attribute to vary according to how well it brings about certain benefits, which in turn can be associated with various needs, which may vary across choice occasions. The propensity to attend is influenced by the gains and costs associated with doing so. The cost comes from the mental effort associated with the evaluation of the attribute. The gain comes from the ability to better distinguish between alternatives. Crucially, the gain is dependent on the link between the attribute and the various benefits, and between the benefits and the needs associated with any choice occasion. The framework is appealing in that it recognises the potential context specificity of preferences for attributes, and the likelihood of not attending to an attribute.

A form of ANA is considered by Sims (2003) and Sims (2010) in the context of macroeconomic behavioural models. Sims treats the individual as having a limited ability to process information, and integrates such a constraint into a dynamic programming problem. This corresponds with the ANA literature somewhat, in that not all information is processed or attended to. The model outputs align well with much observed macroeconomic behaviour. A notable aspect of the market interactions that are modelled is that multiple individuals may rely on the same sources of information, for example, financial information a newspaper. Thus, other agents can influence what information is attended to, and this may induce correlation in information attendance across individuals.

2.2.3 Key issues

The following sections provide a critical analysis of a number issues concerning ANA that are important, and in most cases, prominent in the literature. Many of these issues are key motivations for the research agenda in this thesis.

Stated verses analytically derived attribute nonattendance

The advantages and disadvantages of using stated and analytical methods have been detailed in the preceding sections. This section broadly addresses some of the issues associated with the two key approaches, and the decision as to which to employ.

Perhaps the most prominent line of critical assessment with regards to stated ANA is whether such responses are accurate, and can be relied upon by the analyst. The broadly informed conclusion is that it cannot be treated as if it is completely accurate, where this conclusion has been reached via a triangulation of significant sensitivities for stated ignorers (e.g., Hess and Rose, 2007), inconsistency between stated ANA and conditional parameter estimates (Hess and Hensher, 2010), and the use of stated ANA as a covariate in the LC approach (Hole et al., 2012). The first of these reasons is probably the most compelling, as with the other two, it is not possible to say definitively whether the difference is due to inaccuracy in the stated ANA, or a problem with the analytical method employed. However, as noted in Section 2.2.2, the literature has tended to interpret significant sensitivities for stated ignorers as all such individuals having attended, usually with a milder sensitivity than stated attenders. Instead, some respondents may truly ignore the attribute, while others attend and exhibit some degree of sensitivity.

Balcombe et al. (2011) provided contrary evidence, showing in their study that specifying separate coefficients for stated ignorers did not lead to improved model fit, in three out of four variants of their choice experiment. However, that this was not the case for one variant demonstrates that even in very similar choice contexts, the evidence varies, and stated ANA cannot be clearly judged as superior or inferior. The conclusion to draw is that testing a variety of methods in any given context instills confidence that the method ultimately adopted is appropriate. Applied specifically to the stated verses inferred ANA debate, this would suggest collecting stated ANA, testing methods that do and do not leverage such information, then choosing the method deemed most appropriate. If stated ANA is not collected, such comparisons can never be made. However, stated ANA is not without its costs.

Any additional data points in an SC survey will take extra time to collect, and so come at an additional cost. This could be realised through additional face-to-face interviewer time in a personal interview, extra per survey cost if the sample is recruited via an online panel, or high costs of incentives to provide to the respondent. A longer survey also places additional burden on the respondent, and may decrease their engagement with the survey, compromising response quality. The stated ANA questions should be after the choice tasks, and so are likely to be towards the end of the survey, possibly compromising the quality of the stated ANA responses specifically. Finally, stated ANA methods cannot be used for old datasets for which stated ANA responses were not collected. For these datasets, an analytical method is obviously advantageous.

The analytical techniques, however, are somewhat dependent on the choice model having a panel specification, with multiple responses per individual. The conditional parameter estimate approach of Hess and Hensher (2010) relies on a conditioning on the sequence of choices made by each individual. The greater the number of choices, the more accurate the conditional parameter estimates will be (Train, 2009). Indeed, Mariel et al. (2011) found with simulated data that the accuracy of inferred ANA increased with the panel length. Hole (2011b) found with the LC approach that a panel specification was required: they could not estimate a cross sectional version of their model. More responses per respondent may also increase the cost of data collection, especially if some degree of complexity is introduced into the choice tasks[16]. The number of respondents could be reduced to compensate, however, that may compromise the diversity of the respondents in the sample. Also, a longer panel may be more burdensome to the respondent, resulting in fatigue effects (Bradley and Daly, 1994). A panel specification of a choice model may cause problems if ANA behaviour varies across choice tasks. The impact of choice task ANA on both stated and analytical methods will be discussed in some detail in the next section.

Whereas under the analytical approach, ANA is endogenous, and can easily be applied beyond the sample used to infer ANA, stated ANA cannot be so readily applied. Hensher (2007) provided the example of time savings benefits for road infrastructure, where the ANA incidence rate would have an impact on the benefits across a population. Hensher suggested making some assumptions about the incidence rate of each pattern of ANA in real world applications.

[16]Indeed, complex choice tasks, defined in terms of the dimensions of the choice tasks, are arguably more appealing to the analyst once ANA can be accommodated to reduce any biasing influence. The respondent can attend to the attributes that matter to them. Relevancy is key (Hensher, 2006a).

Hensher and Greene (2010) made the important observation that the analyst does not know whether stated or analytical methods are closer to the 'truth' in terms of determining ANA. Model fit may provide some clues, but should not be the sole criterion. Perhaps the most promising approach is to employ an analytical approach that additionally leverages stated ANA responses. Hole et al. (2012) did this in the context of the LC approach, with a subsequent improvement in model fit. Attribute nonattendance was handled probabilistically, but the probability was allowed to vary according to the stated ANA response, which offered a refinement on the naïve probability. Analytical methods represent an advance on simply constraining marginal utilities to zero based on stated ANA responses. Nonetheless, stated ANA responses may still have much to offer.

Serial verses choice task attribute nonattendance

A number of papers have investigated whether the ANA behaviour of a respondent varies across the choices tasks that they complete, and if so, what the implication may be for model outputs and fit. Scarpa et al. (2010) acknowledged that collecting nonattendance information at the choice task level comes at the cost of survey time, and were motivated to find out if the amount of variation across choice tasks justified the cost.

Puckett and Hensher (2008) and Puckett and Hensher (2009) presented the findings from a study into decision making with regards to road freight. They asked respondents which attributes they ignored after each choice, for each alternative within the choice task. In addition to ANA, Puckett and Hensher (2008) handled the stated aggregation of common-metric attributes, and found that VTTS values decreased in magnitude once both processing strategies were handled in the model. Puckett and Hensher (2009) investigated the patterns of ANA at the choice task level more closely. They found very few differences across the alternatives in each choice task, and only minor differences across the choice tasks for each respondent. Their findings do not appear to recommend capturing ANA at the choice task level, although they called for further research, with more choice tasks, and in different choice contexts.

Puckett and Hensher (2009) and Meyerhoff and Liebe (2009) both reported the stated ANA rates for each of the positions along the panel (i.e., whether presented first, second, etc). Neither observed any discernable pattern. Scarpa et al. (2010) also found that the position has no impact on a model explaining the propensity to not attend to an attribute. However, this type of analysis does not consider what is happening with respect to ANA, over choice tasks, at the level of the individual. Meyerhoff and Liebe (2009) additionally provided a number of measures at the individual level. They considered the number of choice tasks in which a respondent ignored at least one attribute, and found that while 20 percent always ignored at least one attribute, and 50 percent never did, 30 percent attended to all attributes in some tasks, and ignored at least one attribute in others. Further, while eight percent of respondents always ignored the same combination of attributes, 42 percent changed the combination at least once. Scarpa et al. (2010) reconstructed serial ANA from the choice task responses, where this measure is by definition at the level of the individual. They found choice task ANA to be far more prevalent than serial ANA. For example, cost was ignored in 20 percent of choice tasks, but only over all choice tasks for five percent of respondents. Overall, there is evidence in a number of studies that ANA can vary for each respondent over a sequence of choice tasks.

Meyerhoff and Liebe (2009) and Scarpa et al. (2010) both compared the model performance when accounting for serial and choice task ANA. Meyerhoff and Liebe (2009) reconstructed 'serial' ANA by classifying a respondent as serially not attending if they ignored an attribute for at least one choice task, rather than all. They found that a model handling serial ANA outperformed the baseline model with no treatment of ANA, on the AIC, and that both of these models were outperformed by the model that handled choice task ANA. Although no significant difference in WTP was observed, they cautioned against merely considering serial ANA.

In contrast, Scarpa et al. (2010) handled serial ANA in a more plausible way, defining serial ANA as not attending to an attribute across *all* choice tasks. They found the same ordering of model fit as Meyerhoff and Liebe (2009), with the model naïve to ANA performing worst, and the choice task ANA model performing best. Willingness pay measures were observed to change in both directions, although generally the WTP values were lower under choice task ANA than either serial ANA, or the naïve model. The significance of the changes was not reported, however, it was found that the WTP values for the choice task ANA model had better t-ratios, and were of more plausible magnitude. Overall, Scarpa et al. (2010) concluded that obtaining and accounting for choice task ANA was advantageous.

Mariel et al. (2011) tested the conditional parameter estimate approach of Hess and Hensher (2010) on simulated data (see Section 2.2.3 for more details). They found that when ANA varied across choice tasks as well as across respondents, the accuracy with which ANA could be inferred was compromised. Whilst unsurprising, it demonstrated that the accuracy of analytically derived ANA may be compromised by choice task ANA.

A potential problem with collecting information on how the respondent is processing the information after every choice task is that the act of asking them might itself change their behaviour in subsequent choice tasks. For example, they might feel compelled to ignore some of the attributes. Collecting the information at the end overcomes this problem, although it is itself susceptible to problems such as the respondent having difficulty recalling (Scarpa et al., 2010). An analytical method could potentially overcome both concerns, however, existing approaches are largely constrained to treating ANA as being invariant for each respondent over the length of the panel. The conditional parameter approach relies on a single, conditional parameter distribution informed by all choices made by that individual. If censored distributions are employed, then treating the continuous component of utility as independent across choice tasks for the individual is not plausible in most instances, yet it is the continuous distribution that, through the censoring, controls the probability of not attending to an attribute. Thus, relaxing the assumption of serial ANA necessitates relaxing the assumption of preference invariance along the panel, which is likely to be more difficult to support.

The use of analytical methods for inferring ANA raises another issue in the context of choice task ANA. If the methods rely on an assumption that ANA is serial in nature, as with the conditional parameter estimate approach, then the estimated ANA rate might be lower than the true ANA rate. For example, if a respondent attends to an attribute for half of the choice tasks, but ignores it for the other half, then the conditional parameter estimate is likely to be significant, but muted in magnitude from the true sensitivity expressed in the choice tasks in which they attended. Put another way, the conditional parameter distribution is unlikely to have a mean near zero, and a high coefficient of variation. Scarpa et al. (2010) provided an example whereby cost was not attended to in 20 percent of choice tasks, but not attended to over all tasks by only five percent of respondents. Whilst likely dependent on the specifics of the choice task levels, it is reasonable to expect that the inferred ANA rate would be closer to five percent than 20. This hypothesis could be tested with simulations, although it will not be in this body of work, and will remain an area for future research.

Even if analytical methods do capture serial ANA better than choice task ANA, this is arguably a better outcome than some alternatives, such as being able to capture all instances of ANA, but not being able to separate out serial ANA. Serial ANA is a stronger and more generalisable behaviour, that may be of more interest to the analyst. Someone who does not attend across all choice tasks is more likely to be doing so because it represents their preferences (or lack thereof). Conversely, someone who does not attend only in certain choice contexts is more likely to be doing so because of some aspect of the experimental design, or due to some local framing effect. Without dismissing the validity and importance of these alternative reasons for ANA, genuine indifference to an attribute is likely a more useful inference to make from a model, and is more applicable beyond the data on which the model was estimated. Choice task ANA might be better captured analytically by applying some deterministic rule, such as those found in Cameron and DeShazo (2011). Indeed, Scarpa et al. (2010) tested for some of the measures suggested in Cameron and DeShazo (2011), but only found limited supporting evidence.

Attribute nonattendance heterogeneity

Just as the same attribute or attribute level may invoke a range of sensitivities across individuals, or even choice tasks, so too may the same attribute invoke a range of propensities to attend to the attribute, across individuals, and, possibly, choice tasks. If ANA is seen exclusively as a product of the way in which information is integrated, then these varying propensities could be considered as 'process heterogeneity' (Puckett and Hensher, 2009). If ANA reflects true indifference to the attribute, irrespective of context, then these varying propensities could just be one aspect of taste heterogeneity.

Remaining agnostic as to the cause, it will simply be referred to as 'ANA heterogeneity' herein. If both attendance and nonattendance occurs in some choice context, by definition there is ANA heterogeneity: some attend, others do not. The simplest understanding of ANA is at an aggregate level, whereby some percentage of a group of decision makers do not attend to an attribute. This is akin to random ANA heterogeneity. A more nuanced understanding of ANA recognises that the percentage of ANA may vary between some discernable subgroups. That is, there may be systematic sources of ANA heterogeneity. This distinction of random verses systematic ANA heterogeneity parallels the concept of random verses systematic taste heterogeneity. One way to capture systematic ANA is to enter attributes only as interactions with other variables such as socio-demographics. If some of the interactions are not significantly different to zero, and others are, then ANA is varying systematically. The problem here is that the appropriate variables need to be selected, and in any case such variables may not exist.

Attribute nonattendance heterogeneity may be handled in two key ways. The analyst can seek merely to understand what is causing differing propensities to not attend, should such differences exist. This has been typically achieved by estimating a model with ANA, either stated or inferred, as a dependent variable (Hensher, 2006a; Scarpa et al., 2011; Rose et al., 2012a; Carlsson et al., 2010; Scarpa et al., 2010). More ambitiously, the analyst can attempt to include ANA heterogeneity into a choice model with the choice alternative as the dependent variable, which is likely to be their primary model of interest (Hess and Rose, 2007; Hensher et al., 2007; Hole et al., 2012).

Hensher (2006a) investigated the role that choice task dimensionality (distinct in his view from complexity) plays on the propensity of a respondent to state that they ignored attributes[17]. Within the choice tasks, he varied, across respondents, the number of alternatives, number of attributes, number of attributes levels, and the range of the attribute levels relative to a reference alternative that was a recent car trip. An ordered logit model was employed, where the dependent variable was the number of attributes which were ignored. He found that the aggregate ANA rate increased as the number of attribute levels increased, the difference between attribute levels decreased, the number of alternatives decreased, and the range of the attribute levels relative to the reference alternative decreased. This provided clear evidence that ANA heterogeneity may be caused in part by decisions surrounding the design of the choice tasks in an SC study.

Another study that investigated the impact on ANA of some aspect of the construction of the choice task was that of Scarpa et al. (2011). They investigated differences in inferred ANA at each rank of a rank-ordered choice model, that was estimated on data collected with the best/worst (class three) elicitation method. Using a multivariate probit model, they found some gender influences, plus a limited degree of consistency in ANA across the ranks.

The two studies above imply that ANA heterogeneity is not intrinsic to a group of decision makers, but is instead induced by some ex-ante decision by the analyst regarding the dimensions of the choice task (Hensher, 2006a), or is at least somewhat inconsistent across the multiple choices that can be made when three or more alternatives are available (Scarpa et al., 2011). If, however, ANA is an expression of taste heterogeneity, and represents genuine disinterest in the attribute, then the analyst might want to know who is more likely to ignore an attribute and, ideally, why. For example, managers may wish to know who is not attending, so that they can target a marketing campaign (Rose et al., 2012a). Rather than introduce some properties of the choice experiment (e.g., dimensions, rank of choice) as explanatory variables for the incidence of ANA, socio-demographic and other information about the individual can be introduced instead.

Carlsson et al. (2010) estimated a multivariate probit model, with stated ANA as the

[17]Stated ANA was employed, and so it would not be possible to know for certain whether that attribute actually was ignored. The same caveat applies to inferred ANA.

dependent variables. They found only a small number of significant influences, including age and university education, on just some of the attributes. Scarpa et al. (2010) estimated binary logit models for serial and choice task ANA, with stated ANA also the dependent variable. A more extensive array of significant influences was established, including family size, age, income, and reason for visiting a nature park. Interestingly, for choice task ANA, no difference was observed along the panel, between questions asked earlier and later, signifying a lack of learning and fatigue (Bradley and Daly, 1994). Crucially, the model pooled ANA responses for all attributes, and just estimated attribute specific constants, with the consequence that, as specified, drivers of ANA could not be ascertained for each attribute, only for all attributes as a whole.

The other key approach for handling ANA heterogeneity is within the main choice model of interest. As discussed, testing for systematic influences on ANA is somewhat analogous to capturing systematic sources of taste heterogeneity by interacting the attribute with one or more other variables, such as socio-demographic information about each individual. The actual mechanism for achieving this with ANA depends on the type of model employed. Under the LC approach, wherein ANA is inferred, the class assignment probabilities can be parameterised by the covariates of interest. Hess and Rose (2007) found a range of significant socio-demographic influences. As noted previously in Section 2.2.3, Hole et al. (2012) introduced stated ANA as covariates in the class assignment, thus varying the probability of ANA based on the respondents' belief that they ignored an attribute. If censored RP distributions are employed, then an interaction between covariates and either the mean or the variance of the distribution will vary the ANA rate, however, as discussed in Section 2.2.2, there may be confounding between taste heterogeneity and ANA, and introducing covariates only complicates the confounding. For both the LC and conditional parameter estimates approaches, not introducing covariates is equivalent to ANA heterogeneity being purely random.

As detailed in Section 2.2.2, Hensher et al. (2007) stochastically handled stated ANA by additionally considering the influence on ANA of age, income, and the attribute levels of the choice tasks. Puckett and Hensher (2008) examined the stated ANA rates in their study, and found large differences between the two main types of respondent: freight transporters and freight shippers, with ANA rates of up to 12 percent for the former, and 40 percent for the latter. This shows that different types of respondent may have very different ANA behaviour. This is somewhat inconsequential when stated ANA is being employed, as any differences are simply observed from the stated ANA responses. However, unless adequately handled in an analytical method, these differences may not be detected. Segmentation could be employed, with distinct ANA behaviour estimated for each segment. Under the LC approach, the various segments could be entered as dummy variables in the class assignment component of the model. In conclusion, there may be much to be gained from not just trying to explain ANA behaviour via separately estimated models with ANA as a dependent variable, but by integrating ANA heterogeneity into the main choice model of interest.

Impact on willingness to pay measures

Willingness to pay measures are a specific type of marginal rate of substitution, where the substitution is between a choice attribute or alternative, and money[18]. Willingness to pay measures are crucial model outputs for choice models in a wide range of fields, including transportation, marketing, health economics, and environmental economics. Any bias in WTP may result in sub-optimal policy, pricing, and product development decisions. Unsurprisingly then, one of the primary aims of the ANA literature has been to evaluate the impact of ANA on WTP. The first line of enquiry has been an empirical comparison of the magnitudes of the WTP measures before and after ANA is handled. The overwhelming evidence points to the *potential* biasing influence of ANA on WTP. However, no consistent direction in the bias has been found across many studies. The second line of enquiry has been more conceptual in nature, and concerns how to interpret and treat the portion of the sample that ignores one or both of the WTP attributes, with ignorers of cost being particularly problematic. Both lines of enquiry will now be discussed.

Many studies have compared the WTP values derived from a model that assumes all attributes are attended to (the 'naïve model'), with those from a model that handles ANA. The WTP values typically reported in the models that handle ANA are for individuals who attended to both cost and the attribute for which the WTP value is being derived (the 'WTP attribute'). The evidence is decidedly mixed, across a large number of studies. Most common is a decrease in WTP, for most attributes, once ANA is accounted for (Hensher et al., 2005; Rose et al., 2005; Hensher, 2006a; Campbell et al., 2008; Hensher, 2008; Puckett and Hensher, 2008; Campbell and Lorimer, 2009; Scarpa et al., 2009; Puckett and Hensher, 2009; Campbell et al., 2010b; Carlsson et al., 2010; Campbell et al., 2012, 2011; Hole, 2011a; Hole et al., 2012). Numerous studies have observed an increase in WTP for most or all attributes (Hensher, 2007; Hensher et al., 2007; Hensher and Greene, 2010; Hensher et al., 2012a). Others have found any differences to not be statistically significant (Meyerhoff and

[18]Money has been captured through a variety of mechanisms in choice studies, including attributes for product prices, changes in tax paid, fares and tolls.

Liebe, 2009; Scarpa et al., 2011), and Scarpa et al. (2010) found the direction of change to be very varied across attributes. Some studies, but not all, tested for statistical significance of the differences in WTP.

Broadly, the bias in WTP will be dependent on the relative rates of nonattendance to the cost and WTP attributes, since nonattendance to either attribute would bias downwards the magnitudes of the associated taste coefficient(s). For example, Scarpa et al. (2009) noted how nonattendance to cost would pull the cost coefficient downwards, and so push the WTP upwards. In the MNL model, the biasing forces are relatively simple. A single coefficient is estimated each for the WTP attribute (or attribute level), and the cost attribute. These coefficients are biased downwards to some extent by ANA. The WTP decreases if the WTP attribute is biased downwards proportionally more than the cost attribute; if the bias is less than the cost attribute, the WTP increases. Any incidence of ANA in either attribute is likely to induce some bias, although the difference may not be statistically significant. However, for other models, the WTP may be biased in more complex and subtle ways.

For the RPL model, the ANA might impact on both the mean and *variance* of the RP distribution. Hensher (2007) found that the incidence of implausibly signed VTTS reduced once ANA was accounted for[19], and that the range of VTTS values decreased. This second phenomenon was also observed by Hess and Hensher (2010). This suggests that ANA might be captured to some extend in a naïve model by inflated preference heterogeneity. The biasing influence may well be less predictable. Chapter 4 will closely examine the nature of the biasing influence of ANA on a RPL model, using simulations.

The biasing influence may extend to the covariances of the RP distributions. This is because nonattendance to certain combinations of attributes may be captured in the covariances of the preference heterogeneity. For example, a high incidence rate of dual nonattendance to two attributes might induce an excessively high correlation between random parameters introduced for each. The correlation in ANA may be erroneously captured by correlation in taste heterogeneity, due to the inability of the model to adequately separate out ANA. In contrast, correlation could be (and frequently is) captured in stated ANA responses, which are then used constrain the taste coefficients to zero. Alternatively, under the LC approach, each combination of ANA across the attributes can be freely estimated. With so many avenues for biasing the model outputs, it is indeed not surprising that the biasing outcomes observed in the literature[20] are so variable.

[19]See Section 2.2.3 for a more extensive discussion.

[20]The bias is observed, but inherently uncertain, for it is not known if stated or inferred ANA is correct,

The second line of enquiry has been more conceptual in nature, and concerns how to interpret and treat the portion of the sample that ignores one or both of the WTP attributes, with ignorers of cost being particularly problematic. Handling nonattendance to the WTP and cost attributes may result in a number of situations that are awkward for the computation of WTP: when the WTP attribute is ignored, and so has a marginal utility of zero; when the cost attribute is ignored; and when both the WTP and cost attributes are ignored. Rose et al. (2005) note that when a non-cost attribute is ignored, there is a zero WTP for that attribute, and that when the cost attribute is ignored, the WTP becomes infinite, irrespective of whether the WTP attribute is attended to or not.

Looking first at nonattendance to the WTP attribute, a zero WTP is not particularly controversial. If the individual is genuinely indifferent to an attribute, then they will not wish to expend any money to obtain it. Several authors have urged caution in such an interpretation, without ruling it out altogether (Campbell et al., 2008; Balcombe et al., 2011). Campbell et al. (2008) and Hensher et al. (2012a) both noted that the attribute *levels*, rather than the attribute, may have been ignored. Clearly, the motivation for ANA affects the plausibility of the associated zero WTP. Hensher et al. (2012a) highlighted the implausibility of ignoring travel times savings, although Cirillo and Axhausen (2006) argued that it may be possible in some circumstances. The key point here is that a zero WTP for some types of attributes may be implausible.

Nonattendance to cost is far more controversial, and precludes estimation of WTP for the subsample that does not attend. The literature has noted that a marginal utility of money of zero is unrealistic (Scarpa et al., 2009; Balcombe et al., 2011) and Carlsson et al. (2010) additionally noted that a nonzero utility of money, and so disutility of cost, exists, but just cannot be determined from the survey. Carlsson et al. (2010) postulated several causes of cost ANA that may be prevalent in non-market valuation studies, including extreme 'yea-saying', and protests against trading environmental attributes with money. Others have suggested that nonattendance to cost may be an artefact of the choice experiment process or design (Balcombe et al., 2011; Hess et al., 2011), with inappropriate ranges of price frequently cited (Balcombe et al., 2011; Hensher et al., 2012a). Hensher (2006a) found that ANA rates, in aggregate, increased as the range of the attributes decreased. Indeed, cost could plausibly be ignored if it did not vary enough. Scarpa et al. (2009) called for future research into finding ways to make attribute ranges relevant to each individual. They also suggested that the WTP for cost nonattenders should be constructed using the cost sensitivities recovered for

and ANA has been handled appropriately in the model.

cost attenders, arguing that the much more plausible WTP values justifies such a move.

A number of non-market valuation studies have observed very high ANA rates for cost, and a marked decrease in WTP once ANA was accounted for, in some cases by an order of magnitude (Scarpa et al., 2009; Campbell et al., 2010b, 2011, 2012). Crucially, in these studies, the WTP values became much more plausible once ANA was handled. This demonstrates the profound impact cost ANA can have on WTP, both numerically and behaviourally, if full attendance is instead assumed. Hensher (2007) found that the relative VTTS of free flow and slowed down time became more plausible once ANA was accounted for. Scarpa et al. (2010) found that handling ANA at the choice task level resulted in WTP values more in line with prior expectations, relative to serial ANA. Thus, accommodating ANA may not only change the magnitude of WTP, but make the WTP values more plausible in a variety of ways.

Campbell et al. (2012) observed little difference in mean WTP once they handled taste heterogeneity for cost, and ANA for specific attribute levels, rather than the attribute as a whole. However, the mean WTP they computed was informed by much more of the sample, and so more confidence could be placed in it.

The potential impact on WTP of not adequately capturing ANA draws a number of issues into focus. The motivation for each individual not attending to an attribute is important, because if ANA is not due to a true indifference to the attribute, then the WTP might be called in to question, and the transferability of the WTP values beyond the choice tasks presented in an SC study may be limited. There appear to be particular problems with WTP in non-market valuation studies.

Also important is the accuracy with which ANA is retrieved, either through stated responses, or analytical inference. If the econometric method employed to infer ANA was found to induce its own bias, then the accuracy of the WTP measures would be compromised. Thus, WTP is a strong motivator to find a robust econometric model that can infer ANA without bias.

As Hensher et al. (2012a) noted, problems with WTP arising from ANA are probably common, but are not frequently tested for. The existence, direction and magnitude of the bias will depend on the dataset, but the potential exists for the bias to be profound, and so testing for ANA is important. The analyst should critically evaluate whether the ANA rate, and by extension the WTP, is reasonable in any given context.

Systematic investigations of the impact of attribute nonattendance

Many studies have examined the impact of ANA on the parameter estimates and WTP measures, by comparing the values before and after ANA is accounted for (e.g., Rose et al., 2005; Hess and Rose, 2007; Campbell et al., 2008). If we had complete confidence in stated ANA responses, or the performance of the analytical methods, then the bias introduced when ANA is unaccounted for could be quantified with confidence. However, we can have no such confidence in either stated or inferred ANA, as these techniques may introduce their own biases (Hoyos et al., 2010). Consequently, the true ANA rate cannot be determined. Alternatively, if simulated datasets are used, then the true parameter estimates and ANA rates will be known, and the degree of bias can be determined. This also provides a way to evaluate the potential accuracy of any given technique for recovering ANA, although further problems may be encountered when the technique is applied in an empirical context. Two studies have examined bias and model performance with simulated data in the context of ANA (Hoyos et al., 2010; Mariel et al., 2011), and are detailed below. Simulated data have also been used to investigate the impact of other decision rules. For example, Rose et al. (2012b) considered lexicographic choice, inconsistent choice, non-trading from a reference alternative, and changing sensitivities over the length of the panel.

Hoyos et al. (2010) examined the reliability of the stated ANA methodology, and the LC approach, using simulated datasets with various ANA rates. The simulated choice tasks contained three alternatives and four attributes, each described by five levels. One attribute was price, and fixed coefficients were specified such that the WTP for the remaining attributes was one. Attribute nonattendance was specified for two of the non-price attributes, at rates of 20, 40 and 60 percent. Consequently, bias was considered in terms of WTP, with ANA impacting on the non-price attribute only. Each dataset contained 1600 responses, and 1000 datasets were generated for each treatment. Four main investigations were performed. In the first, there was no correlation in the errors of the utility functions. Then, correlation was introduced into the error terms, across sets of four choice tasks, representing the four choices of a respondent. The third investigation correlated the errors across the two non status quo alternatives, while the fourth investigation simultaneously introduced both forms of correlation. Stated ANA was tested by setting the coefficients of ignored attributes to zero, with full accuracy. The LC approach was tested for one attribute only, with the coefficient for that attribute set to zero in one of two classes.

First, it was found that if ANA was not handled, the magnitude of the WTP decreased by approximately the same percentage as the ANA rate. With no correlation in the errors, the WTP values were recovered with no bias. However, the LC model either over or underestimated the WTP values and ANA rates, depending on the true ANA rate. When the errors were correlated across choice tasks, bias was introduced under both approaches. When correlation was introduced across the non status quo alternatives, stated ANA was once again accurate, but the LC model tended to underestimate WTP and overestimate the ANA rates. That stated ANA was mostly reliable is not surprising, as it was assumed that stated ANA was accurate. What is contentious about the use of stated ANA, however, is whether the responses themselves are accurate, where inaccuracy would likely lead to bias in the WTP values, depending on the relative incidences of false positives and negatives. The LC approach was generally quite unreliable, which is the important finding of the paper. Note that preference heterogeneity was not accounted for by Hoyos et al. (2010). This thesis will also employ simulated data to investigate ANA, however, in all cases the simulated respondents will exhibit preference heterogeneity, which lends more behavioural realism, and allows the interrelation of true preference heterogeneity and ANA to be investigated.

Mariel et al. (2011) extended upon the work of Hoyos et al. (2010) by specifying a distribution of sensitivities to one of the attributes in a range of simulated datasets, varying the nonattendance rate to this attribute, and investigating the performance of the conditional parameter estimate approach of Hess and Hensher (2010) on these datasets. A variety of treatments were specified. Two panel lengths were tested, of five and fifteen respondents. Attribute nonattendance, at rates of 20, 40 and 60 percent, was specified both serially, across all of an individual's choice tasks, and at the choice task level.

First, the biasing impact of ANA on the estimated coefficients was tested, with the conventional RPL model. Under serial ANA, bias was only detected on the attribute for which ANA was imposed. Under choice task ANA, bias was additionally observed on the other attributes, which had no ANA. As the true ANA rate increased, so too did the downward bias in the mean of the estimated random parameter. Bias was evident on the standard deviation, but was upward for serial ANA, and downward for choice task ANA. Second, setting the coefficients to zero for ignored attributes removed all bias, but this is unsurprising, and again, the true concern with the stated ANA approach in empirical applications is whether the stated responses actually match the choice behaviour.

Third, and more revealing, were the findings concerning the conditional parameter estimate technique of Hess and Hensher (2010). Recall from Section 2.2.2 that this approach relies upon the specification of a threshold coefficient of variation value. Mariel et al. (2011)

evaluated the performance of this approach across a range of thresholds, using the percentage of respondents correctly identified as being nonattenders. They found that as the panel length increased, so too did the accuracy of the inference of ANA. The worst performance was observed with choice task ANA, which necessitated the estimation of a cross sectional RPL model. Crucially though, the best performing threshold was different for different true rates of ANA. This is problematic, as while the analyst might have some sense of what the true ANA rate might be, it is the very thing they are trying to estimate. Encouragingly, Mariel et al. (2011) noted that the parameter estimates recovered were unbiased, but details were not provided in the paper.

Mariel et al. (2011) demonstrated how simulations are a useful way to reveal the biasing influence of ANA, and evaluate the performance of various techniques for handling ANA. However, they make too strong a conclusion that ANA "can be inferred with certain guarantee" (p.29), and that the inferred ANA can be used to determine whether respondents do what they say. Simulations provide a very controlled environment in which to test a model. In an empirical application, there is no such control, and there may be a whole array of influences that could undermine the performance of a model and the confidence that can be placed in its outputs. For example, different SC experimental designs may have a very different power to extract conditional parameter estimates, and the ANA condition may be inferred erroneously as a consequence. Simulations can provide a certain level of confidence in a model or technique, and are particularly useful for identifying fundamental problems, but further problems may arise in more realistic, empirical choice contexts. Testing in these empirical contexts is important also, and conclusions drawn from simulations must be tempered.

Simulations will be used in this thesis, in Chapter 4, to further investigate the influences of ANA on model outputs. As with Mariel et al. (2011), a distribution of sensitivities will be specified when generating the datasets, and RPL models will be estimated. However, a wider range of distributions will be tested, and the standard deviations or spreads of these distributions will be systematically varied. This is motivated by the belief that any mass of coefficients *near* zero is likely to have some impact on the extent of bias induced by ANA, and that such a mass may also impact on the ability of a model to estimate ANA analytically. Whereas Hoyos et al. (2010) investigated the performance of the LC approach when the true sensitivities were the same across respondents, this body of work will investigate the performance in a more realistic context wherein those true sensitivities vary across respondents. Additionally, the performance of the proposed approach, the RPANA model, will be tested with simulations in Chapter 5, prior to application in an empirical context in Chapter 6.

Incidence of implausibly signed random parameter coefficients

Many of the distributions commonly employed to represent preference heterogeneity in the RPL model are not bounded with respect to the sign of the coefficients generated. That is, the distribution may have support in both the positive and negative domain, and indeed, unbounded distributions such as the normal always will. Other distributions, such as the uniform and triangular, are bounded, but may still span zero and generate coefficients of both signs. However, in many contexts, coefficients and WTP values of a certain sign will not be behaviourally plausible. Examples include positive coefficients for costs, fares, and travel time[21], and negative coefficients for legroom on aircraft and the battery life of a mobile phone. At most, an individual may be indifferent to these attributes or features (Train and Sonnier, 2005), which of course is relevant in the context of this body of work. The case whereby some coefficients are not behaviourally plausible will be referred to herein as 'sign violation'.

Sign violations can and frequently are overcome by specifying a distribution that enforces a single sign. The lognormal distribution was used in some of the earlier random parameter choice models (Ben-Akiva et al., 1993; Brownstone and Train, 1998; Bhat, 1998), and has been used extensively since. Other options include the Rayleigh (Hensher, 2006c), and distributions that place a constraint on the variance, including the constrained uniform and triangular distributions (Train, 2001; Hensher and Greene, 2003). However, Hensher (2006c) noted that the effectiveness of such constraints in enforcing sign might be undermined by systematic heterogeneity in either the mean or the variance. He proposed a solution to overcome this problem. More problematic is that simply imposing a constraint may mask the true cause of the sign violation, and result in a misspecified model. For example, it is obvious that ANA is likely to have a downward bias on the magnitude of the mean of a distribution of sensitivities to the attribute. It is possible also that ANA could impact upon the dispersion of the distribution, which, for distributions that are unconstrained in sign, could increase the incidence of implausibly signed coefficients. Hensher (2007) made such a finding.

In his empirical study of alternative packages of time and cost for car drivers, Hensher (2007) obtained stated ANA rates ranging from 6.3 to 27.5 percent. Value of travel time savings were calculated based on conditional parameter estimates, both with ANA accounted

[21] Although travel time has been debated; see for example Mokhtarian and Salomon (2001) and Cirillo and Axhausen (2006).

for and without. Once ANA was accounted for, across various attributes and distributions, ANA rates dropped from 2.89 percent before to 0.7 percent after, from 5.1 to 0.76 percent, and from 2.3 to 0.64 percent. This suggests that ANA might be one cause of implausibly signed coefficients. Additionally, Hensher (2007) found that, in most cases, accounting for ANA narrowed the ranges of conditional VTTS values, where a similar finding was observed by Hess and Hensher (2010). This suggests that some of what is typically interpreted as preference heterogeneity in RPL models may in fact be due to ANA. It can be argued that ANA is just one form of preference heterogeneity, however, a continuous distribution is not an appropriate way to capture such heterogeneity. A point mass at zero, in addition to a continuous distribution, is more appropriate. Cirillo and Axhausen (2006) proposed and visualised, but did not estimate, such a distribution. Finally, Hensher (2007) obtained more plausible VTTS values after ANA was accounted for, specifically with regard to the relative VTTS for free flow and slowed down time.

Other studies have also noted a reduction in sign violation once ANA is accounted for. Using normal distributions, Train and Sonnier (2005) found that implausibly signed coefficients represented up to 40 percent of the mass of coefficients, including 22 percent for price. Capturing ANA through a censoring of the normal distributions led to a large improvement in model fit. Campbell et al. (2010a) observed a moderate share of coefficients in the positive domain for a normally distributed cost attribute. This share dropped notably when a mixture of normals was employed instead, where one of the normal distributions in the mixture approximated ANA and very low cost sensitivities.

The evidence linking ANA and implausibly signed coefficients appears strong. The following discussion considers this interrelation more abstractly. If a point mass exists at zero, alongside some other distribution of sensitivities, then the ability of the distribution to represent the point mass will depend on the flexibility of the distribution. Most commonly employed distributions have limited flexibility, and so some part of the distribution will serve as a poor approximation to the point mass. There is no reason to expect a coefficient of implausible sign to represent a point mass at zero any worse than a coefficient of plausible sign and equal magnitude. The only advantage of the latter is that it may serve the *additional* function of representing plausible sensitivities. Thus, we can expect some share of implausibly signed coefficients serving to *approximate* ANA. This may be muted somewhat if the distribution is not symmetric, as skewness may move some mass closer to zero, contributing to the approximation of ANA.

If ANA is contributing to implausibly signed coefficients, then reflexively employing a distribution that constrains the sign may bias the distribution, and lead to a misspecified model. A better approach is to first treat the ANA in some appropriate way, and then apply a distribution that constrains the sign, if still required. This approach bears some similarity to the recommendation of Hensher (2006c), which is to first estimate unconstrained distributions and identify the prevalence of implausibly signed coefficients, then account for systematic sources of preference heterogeneity, then finally estimate a model with systematic variation and a sign constrained distribution. In this context, in place of systematic preference heterogeneity, ANA is accounted for.

This thesis will employ the aforementioned simulations to, additionally, gain a deeper understanding of the influence of ANA on sign violation. The simulations will systematically vary the shape of the true distribution, its variance, and the ANA rate, and so determine the degree of sign violation that is induced by ANA, and any interaction ANA has with the true variance of the distribution. Of course, this is in a highly controlled environment. Empirical applications are less controlled, and do not allow a comparison to be made with the true values. Nonetheless, the simulations will provide the reader with a sense of how severe the sign violation may be.

2.3 Contribution to the literature

Section 1.3 has already summarised the contribution of this thesis to the literature. These contributions will be reiterated here, and expanded upon, since the contributions will be more meaningful given the review of the literature in this chapter. Overall, there are two broad areas in which a contribution is made. First, a better understanding is gained into the impact of ANA on model outputs, and the performance of some existing techniques for accommodating ANA. Second, a new method for handling ANA is introduced that overcomes some of the shortcomings of the existing methods. This method will be introduced in general terms here, before being formalised in Chapter 3.

Using simulations, the impact of ANA on the parameter estimates and the incidence of implausibly signed coefficients is systematically investigated. Whilst biasing influences and increased sign violations have been observed in various empirical contexts, the true values were not known, and only a limited number of data points were examined. The simulations allow a more detailed investigation, as the true values are known, and a large number of data points can be tested. Hoyos et al. (2010) investigated the impact of ANA with simulations, but the influence of taste heterogeneity was not tested. Mariel et al. (2011)

did include taste heterogeneity in the simulated data in their study of ANA, but did not vary its extent. The simulations in this thesis will systematically vary the ANA rate, the extent of taste heterogeneity, and the distribution of tastes. Thus, the interaction of these three variables will be considered when examining the impact of ANA on parameter bias and sign violation. Indeed, it will be shown that the true mass of coefficients near zero, due to preference heterogeneity, will impact on the bias that ANA induces. It will also become clear that continuous RP distributions struggle to cope with point masses, such as those representing ANA, and that part of the distribution approximates that point mass. The simulations will provide some sense of the nature and extent of misspecification that results when ANA is not appropriately handled.

The simulations will also be used to evaluate the performance of two existing methods for handling ANA. Hoyos et al. (2010) tested the LC approach with simulations, but did not vary the true taste heterogeneity. By varying the taste heterogeneity herein, it will be shown that the degree of taste heterogeneity has a profound impact on the accuracy of the LC approach. Also, the ability of the censored normal distribution to handle ANA will be investigated with simulations, where it is believed that this is a first.

This thesis also introduces a new technique for handling ANA, the random parameters attribute nonattendance model, which extends upon the LC approach. Latent classes are still employed to constrain sensitivities to zero, however, sensitivities that are not censored are randomly distributed, and so the RPANA model can handle both ANA and taste heterogeneity. This model responds to several calls in the literature. Hole (2011a) suggested that random parameters could be employed within the LC approach to modelling ANA, to handle taste heterogeneity, but did not recognise the problems associated with failing to handle taste heterogeneity. Campbell et al. (2012) did recognise these problems, and reduced them by estimating several nonzero taste coefficients. They called for research into models that can accommodate heterogeneity in tastes as well as processing, which the RPANA model can achieve. A similar model has been developed by Hess et al. (2011), but the proposed RPANA model is more flexible. Differences between the two models, and research agendas, will be noted shortly.

The RPANA model possesses a number of properties which differentiate it from other approaches to handling ANA, and which commend its usage. Unlike the conventional LC approach, the RPANA model can handle taste heterogeneity. The model infers ANA analytically. It does not rely on stated ANA responses, and so is not compromised by any inaccuracy in such responses, or reliant on their collection. Recall that Section 2.2.3 extensively discussed the differences between stated and inferred ANA, and the merits of the latter.

The proposed model may be specified in a parsimonious way, but unlike comparable models, this parsimony can be eroded in a granular fashion, if the assumptions on which the parsimonious specification rely do not hold. There are currently two approaches to the specification of the latent classes in the LC approach to handling ANA. The most common approach estimates one parameter for every combination of ANA across the attributes, and so the number of parameters required increases exponentially as the number of attributes increases. Hole (2011a) proposed an alternative approach, wherein the number of parameters required rises linearly, although this relies on the assumption that ANA is independent between all attributes. The proposed RPANA model allows for various specifications between these two extremes. This allows the independence assumption to be selectively utilised, where it can be supported, to maximise parsimony. The alternative specifications are explored extensively in an empirical context, as are the consequences of making the assumption where it cannot be sustained.

The RPANA model allows covariates to be entered as predictors for ANA. This was achieved by Hess and Rose (2007) in the LC approach, using socio-demographic information. Hole et al. (2012) used stated ANA as a covariate in the LC approach. However, given the limitations of that approach, which will be clearly shown with simulations in Chapter 4, it is not clear whether the stated ANA covariate is capturing ANA, or low sensitivities. The RPANA model reduces the confounding that may occur between ANA and taste heterogeneity, and stated ANA will be tested as a covariate in the model, thus improving upon the methodology of Hole et al. (2012). Note that the RPANA model with covariates has the choice of alternative as the dependent variable, as opposed to a number of studies that have investigated ANA, wherein the choice of stated ANA response is typically the dependent variable (Hensher, 2006a; Scarpa et al., 2011; Rose et al., 2012a; Carlsson et al., 2010; Scarpa et al., 2010). In terms of leveraging covariates for ANA, the RPANA model is an improvement over the censored normal distribution, as any covariate entered into that distribution will influence not just the ANA rate, but also the mean and/or the standard deviation of the distribution.

Unlike the Hess and Hensher (2010) approach that utilises conditional parameter estimates, no arbitrary thresholds need to be specified with the RPANA model. Further, all model specification decisions have a behavioural motivation. The decision unique to the RPANA model, which is the structure of the LC component of the model, is associated with the independence or otherwise of ANA across attributes. Also, unlike the final model in

Hess and Hensher (2010), estimation is simultaneous, with ANA and other model outputs estimated simultaneously. In contrast to the stochastic attribute selection model (Gilbride et al., 2006), which relies on Bayesian inference, RPANA the model can be estimated using classical econometric techniques.

The RPANA model is similar in construction and intent to the model proposed by Hess et al. (2011), however, it is more flexible, and this thesis tests the model in more depth. Note that the research in this thesis was conducted independently of and concurrently with Hess et al. (2011), and the model was developed, implemented and tested before their working paper was made available. Despite the overlap, there are key differences in the models and the research undertaken. In comparing the paper by Hess et al. (2011) and this thesis, the terms 'they' and 'we' will be used, respectively. They employed the latent class structure of Hole (2011a)[22], and thus relied on the assumption of independence of ANA across attributes. As previously noted, this thesis allows for a latent class structure that can selectively relax the assumption of independence of ANA. They did question whether the independence assumption is justified broadly; we will demonstrate that it is not. Where an attribute that is dummy or effects coded is ignored, we censor all coefficients associated with the attribute, where censoring is required, in line with the recommendations of Scarpa et al. (2009). They modelled attendance to each dummy coded attribute level separately, which captured some alternative expression of taste heterogeneity for that attribute, or possibly attribute level threshold effects, rather than ANA. Whereas they only employed the lognormal distribution, we test several other distributions, come to the important finding that certain distributions and variable codings cannot be employed as they result in identification problems, and suggest how the chance of such problems occurring can be minimised. Also, the consistency of distributions between the RPL and RPANA models will be explored. They noted that different distributions will likely have some impact on the estimated ANA rate; we will show this to be the case. Within this body of work, a systematic comparison of the various distributions in an empirical setting, for both the RPL and RPANA model, provides insights into the RPANA model, and what RP distributions may be capturing in RPL models. Another point of difference is the use in this thesis of covariates to vary the ANA rate in the RPANA model, and in particular the use of stated ANA as a covariate. Also, whereas they only tested their model on empirical data, we additionally test the RPANA model on simulated data, to gauge performance in a controlled environment, and gain additional insights into the nature

[22]The version of the Hess et al. (2011) working paper available in August 2012 cited Hole (2011a), but not with respect to this crucial aspect of their model.

of the model.

Thus concludes the outline of the contributions of this thesis. The RPANA model has been introduced here in broad terms. The next chapter will formally define the model.

Chapter 3

Methodology

3.1 Introduction

The model presented in this chapter generalises and extends the LC approach to modelling ANA (Hess and Rose, 2007; Hole, 2011a). Initially, in Section 3.2, fixed coefficients are retained, in what is referred to as the ANA model. Two existing approaches are first broadly outlined, and then a hybrid model is introduced in detail. Since the two existing approaches are special cases of the hybrid model, the latter will be used to precisely define the former. Next, in Section 3.3, the RPANA model is introduced as an extension of the ANA model. This model combines the LC approach for capturing ANA with the use of random parameters for representing preference heterogeneity, conditional on attendance to an attribute. Then, Section 3.3.1 details how the MNL, RPL and ANA models nest within the RPANA model. Section 3.3.2 raises a number of potential identification issues with the RPANA model, and details how these can be avoided. Finally, Section 3.4 compares on a number of dimensions the key papers that employ some form of (latent class) ANA model, and positions the RPANA model in this literature.

3.2 The attribute nonattendance (ANA) model

Consider a choice task wherein the alternatives are described by K attributes. The analyst wishes to model nonattendance to $K^\star$ of these attributes, which may represent all attributes ($K^\star = K$), or a lesser number ($1 \leq K^\star < K$). Choice of a lesser number may be behaviourally motivated, if some attributes are assumed to always be attended to, or econometrically motivated, to lessen the number of parameters that must be estimated.

Under the LC approach, the unconditional probability of respondent n choosing an alternative (or sequence of alternatives across multiple choice tasks) can be decomposed into the probability of that respondent exhibiting a certain pattern of attendance and nonattendance across attributes, and the probability of choosing the alternative or sequence of alternatives, conditional on belonging to a specific class of ANA behaviour. These two components are described in more detail:

Final ANA assignment probabilities These are the probabilities of the respondent exhibiting specific combinations of attendance and nonattendance over $K^\star$ attributes, where there are up to $2^{K^\star}$ possible combinations. Each combination is represented by a class in the LC model. Define M as the set of all realised classes, where $|M| = 2^{K^\star}$ if all possible ANA combinations are to be modelled, or $1 < |M| < 2^{K^\star}$ if some specific ANA combinations are to be omitted. The probability of each respondent n belonging to class m is denoted P_{nm}, and will be referred to as the ANA assignment probability, in recognition of the behavioural interpretation of each class[1]. In most cases herein, P_{nm} will be referred to as the final ANA assignment probability, since this probability may be a function of two or more further probabilities, each of which also controls ANA assignment in some way. Alternate methods for generating P_{nm} will be detailed below.

Choice probabilities conditional on final ANA assignment These are the probabilities of choosing an alternative, or sequence of alternatives across multiple choice tasks, conditional on assignment to a specific combination of ANA. Most examples in the literature employ an MNL model to calculate these probabilities. The choice alternatives are described by K attributes, and nonattendance is modelled for $K^\star$ of these. For each ANA assignment class m, a unique combination of the taste coefficients associated with the $K^\star$ attributes will be constrained to zero, to reflect the specific combination of ANA that the class represents. When not constrained to zero, these coefficients are either constrained to be equal across classes (Scarpa et al., 2009), or unique coefficients are estimated for each class (Hensher and Greene, 2010). The former approach is the most common in the literature. While it requires less parameters to be estimated, it does not capture preference heterogeneity amongst those who attend to the attribute. The latter approach can capture preference heterogeneity which is systematically associated with the ANA pattern imposed.

The unconditional probabilities can be obtained, for each final ANA assignment class m, by multiplying the final ANA assignment probabilities by the choice probabilities that are

[1]This is distinct from the conventional LC model, which has no such behavioural interpretation, with each class merely representing some combination of preference weights for the attributes of the choice alternatives.

conditioned on the ANA assignment, and integrating over the $|M|$ ANA assignment classes.

The most common approach in the literature for generating the final ANA assignment probabilities is to use the conventional LC approach, with a single MNL model employed to calculate each of the $|M|$ ANA assignment probabilities (Hess and Rose, 2007; Scarpa et al., 2009; Hensher and Greene, 2010; Campbell et al., 2011). If all combinations of ANA across $K^\star$ attributes are to be modelled, then the number of parameters required for ANA assignment increases exponentially as $K^\star$ increases. For even a trivial value of $K^\star$, the number of parameters might be prohibitive. However, specific ANA combinations may be omitted at the discretion of the analyst (Scarpa et al., 2009). This decision may be based either on an assumption that the combination of ANA is unreasonable or unlikely, or ex-post evidence that the combination does not occur. This second motivation will be discussed later in the chapter.

An alternative, more parsimonious approach for generating the final ANA assignment probabilities has been proposed by Hole (2011a). Whereas the conventional approach estimates a single MNL model that generates the probability of each combination of ANA across the $K^\star$ attributes, this approach estimates a binary logit model for each of the $K^\star$ attributes, each of which generates the probability of whether a single attribute is attended to or not. These will be referred to as ANA assignment probabilities, as distinct from the final ANA assignment probabilities, which are the probabilities of *combinations* of ANA across the $K^\star$ attributes. The final ANA assignment probability for each ANA combination, P_{nm}, is then the product of $K^\star$ ANA assignment probabilities, each obtained from the binary logit models. The selection of probability (attendance or nonattendance) to include in each element of this product is informed by whether m represents attendance or nonattendance to the attribute in question. There are $2^{K^\star}$ classes in the final ANA assignment model, but as few as $K^\star$ parameters controlling the assignment[2]. However, such parsimony relies on the assumption that the probability of not attending to any one of the $K^\star$ attributes is independent of the nonattendance probabilities of each of the other attributes. If the assumption holds, then the ANA assignment can be estimated more parsimoniously, and the conventional LC approach will be an overparameterisation. If, however, some combination of attributes has a disproportionately high or low probability, then the independence assumption does not hold, and the approach might result in biased parameter estimates and a poorer model fit than the

[2]More than $K^\star$ parameters may control the ANA assignment, if covariates are introduced into the binary logit models. The fully notated generalised model presented below will allow covariates to influence ANA assignment.

conventional LC approach.

Whether the ANA probabilities are independent is likely to vary from one empirical context to the next. Currently, the analyst could test both specifications, and see which best fits the data, using a measure such as the AIC[3]. However, these two approaches represent two extremes of what can actually be considered a continuum. The approach proposed by Hole assumes that nonattendance is independent across all combinations of attributes. The conventional LC approach makes no such assumption, and can handle any correlation structure over the $K^\star$ attributes. The conventional LC approach can replicate the final ANA assignment probabilities obtained under the approach proposed by Hole, however, it does so at the cost of more parameters, where these parameters may be superfluous, if independence holds. Crucially to the development of the generalised approach, it may be that the independence assumption is violated within some subsets of the $K^\star$ attributes, but not between these subsets. The most appropriate model then would be some intermediate point between the two extremes. Such a generalised model is now introduced.

Rather than have $K^\star$ ANA assignment models, each with two classes (Hole, 2011a), or a single ANA assignment model, with up to $2^{K^\star}$ classes (Hess and Rose, 2007), we may have A ANA assignment models, with $1 \leq A \leq K^\star$. Each ANA assignment model a controls the nonattendance associated with $K^\star_a$ attributes. If all combinations of nonattendance to the $K^\star_a$ attributes are to be modelled, ANA assignment model a will have $2^{K^\star_a}$ classes. Specific combinations of attendance can be excluded by the analyst, resulting in fewer classes.

Define C_a as the set of realised classes for each ANA assignment model a, representing all combinations of attendance retained by the analyst. The cardinality of each ANA assignment model, $|C_a|$, will impact upon the cardinality of M, the set of all classes in the *final* ANA assignment model. Specifically, $|M| = \prod_{a=1}^{A} |C_a|$, since all combinations of each realised class in each ANA assignment model will be handled. The probability P_{nc_a}, of respondent n being assigned to class c_a is calculated with an MNL model, such that

$$P_{nc_a} = \frac{e^{(\gamma_{c_a} + \theta_{nc_a} z_n)}}{\sum_{d \in C_a} e^{(\gamma_d + \theta_{nd} z_n)}}. \tag{3.1}$$

A parameter, γ_{c_a}, serves as a constant term, capturing the assignment to class c_a that cannot be explained by other factors. A vector of parameters, θ_{nc_a}, captures socio-demographic and other influences on the assignment of respondent n to class c_a, and z_n is a vector of these influences. To ensure identification, γ_{c_a} and θ_{nc_a} are constrained to zero for one class. Given that most of the discussion in the literature is around attribute *non*attendance, constraining

[3] A likelihood ratio test is not possible, since the two models are not nested.

γ_{c_a} and θ_{nc_a} to zero for the class that represents full attendance to the attributes is the most convenient such constraint to impose. It is also likely that in many empirical contexts, full attendance across the $K_a^\star$ attributes will have the highest probability of all possible ANA combinations, although this is not necessarily the case (e.g., Hensher et al., 2012a).

Recall that each class $c_a \in C_a$ represents a unique pattern of ANA over the $K_a^\star$ attributes which have their ANA state determined by ANA assignment model a. In the final ANA assignment model, each class $m \in M$ will represent a unique pattern of ANA over all $K^\star$ attributes for which ANA is modelled. Now, for some arbitrary class m in the final ANA assignment model, associated with some arbitrary set of ANA assignment model classes, $\{c_1, \ldots, c_A\}$, the probability of respondent n belonging to class m is

$$P_{nm} = P_{n\{c_1,\ldots,c_A\}} = \prod_{a=1}^{A} P_{nc_a}. \tag{3.2}$$

Substituting in Equation 3.1, this becomes

$$P_{nm} = P_{n\{c_1,\ldots,c_A\}} = \prod_{a=1}^{A} \frac{e^{(\gamma_{c_a}+\theta_{nc_a}z_n)}}{\sum_{d\in C_a} e^{(\gamma_d+\theta_{nd}z_n)}}. \tag{3.3}$$

To make the ANA assignment component of the model concrete, three specific examples will be provided: the conventional LC approach, the more parsimonious approach of Hole (2011a), and a third example that lies in between. An analyst is conducting an SC study into airline ticket choice, where the choice alternatives are described by four attributes: fare, flight time (referred to as Time), departure time (Depart), and airline. The analyst assumes that fare will be attended to by all, and thus models ANA for the last three attributes only, hence $K = 4$ and $K^\star = 3$.

Example 1 Under the conventional LC approach, there is one ANA assignment model ($A = 1$). The analyst retains all combinations of ANA, thus $|C_1| = 2^{K^\star} = 2^3 = 8$. Further,

$$\begin{aligned} C_1 = \{&TimeAttend.DepartAttend.AirlineAttend,\\ &TimeAttend.DepartAttend.AirlineIgnore,\\ &TimeAttend.DepartIgnore.AirlineAttend,\\ &TimeAttend.DepartIgnore.AirlineIgnore,\\ &TimeIgnore.DepartAttend.AirlineAttend,\\ &TimeIgnore.DepartAttend.AirlineIgnore,\\ &TimeIgnore.DepartIgnore.AirlineAttend,\\ &TimeIgnore.DepartIgnore.AirlineIgnore\} \end{aligned}$$

The set of all final ANA classes, M, is identical to C_1, thus $|M| = 8$, and

$$\begin{aligned}
M = \{&TimeAttend.DepartAttend.AirlineAttend,\\
&TimeAttend.DepartAttend.AirlineIgnore,\\
&TimeAttend.DepartIgnore.AirlineAttend,\\
&TimeAttend.DepartIgnore.AirlineIgnore,\\
&TimeIgnore.DepartAttend.AirlineAttend,\\
&TimeIgnore.DepartAttend.AirlineIgnore,\\
&TimeIgnore.DepartIgnore.AirlineAttend,\\
&TimeIgnore.DepartIgnore.AirlineIgnore\}
\end{aligned}$$

An example of some potential ANA assignment probabilities is provided below. Since there is only one ANA assignment model, no transformation of the probabilities is required.

$$\begin{aligned}
&P_{TimeAttend.DepartAttend.AirlineAttend} = 0.378,\\
&P_{TimeAttend.DepartAttend.AirlineIgnore} = 0.252,\\
&P_{TimeAttend.DepartIgnore.AirlineAttend} = 0.162,\\
&P_{TimeAttend.DepartIgnore.AirlineIgnore} = 0.108,\\
&P_{TimeIgnore.DepartAttend.AirlineAttend} = 0.042,\\
&P_{TimeIgnore.DepartAttend.AirlineIgnore} = 0.028,\\
&P_{TimeIgnore.DepartIgnore.AirlineAttend} = 0.018,\\
&P_{TimeIgnore.DepartIgnore.AirlineIgnore} = 0.012.
\end{aligned}$$

Example 2 Consider next the approach of Hole (2011a), where $A = K^\star = 3$, $|C_1| = |C_2| = |C_3| = 2$ and

$$\begin{aligned}
C_1 = \{&TimeAttend,\\
&TimeIgnore\}\\
C_2 = \{&DepartAttend,\\
&DepartIgnore\}\\
C_3 = \{&AirlineAttend,\\
&AirlineIgnore\}
\end{aligned}$$

The set of all classes in the final ANA assignment model, M, remains as before.

$$\begin{aligned}M = \{&TimeAttend.DepartAttend.AirlineAttend,\\ &TimeAttend.DepartAttend.AirlineIgnore,\\ &TimeAttend.DepartIgnore.AirlineAttend,\\ &TimeAttend.DepartIgnore.AirlineIgnore,\\ &TimeIgnore.DepartAttend.AirlineAttend,\\ &TimeIgnore.DepartAttend.AirlineIgnore,\\ &TimeIgnore.DepartIgnore.AirlineAttend,\\ &TimeIgnore.DepartIgnore.AirlineIgnore\}\end{aligned}$$

To illustrate the transformation of the probabilities, potential probabilities of attendance and nonattendance are provided for each attribute.

$$\begin{aligned}P_{TimeAttend} &= 0.9,\\ P_{TimeIgnore} &= 0.1,\\ P_{DepartAttend} &= 0.7,\\ P_{DepartIgnore} &= 0.3,\\ P_{AirlineAttend} &= 0.6,\\ P_{AirlineIgnore} &= 0.4.\end{aligned}$$

Then, the probabilities of each class in the final ANA assignment model are products of the appropriate binary attendance probabilities for each attribute.

$$\begin{aligned}P_{TimeAttend.DepartAttend.AirlineAttend} &= 0.9 \times 0.7 \times 0.6 = 0.378,\\ P_{TimeAttend.DepartAttend.AirlineIgnore} &= 0.9 \times 0.7 \times 0.4 = 0.252,\\ P_{TimeAttend.DepartIgnore.AirlineAttend} &= 0.9 \times 0.3 \times 0.6 = 0.162,\\ P_{TimeAttend.DepartIgnore.AirlineIgnore} &= 0.9 \times 0.3 \times 0.4 = 0.108,\\ P_{TimeIgnore.DepartAttend.AirlineAttend} &= 0.1 \times 0.7 \times 0.6 = 0.042,\\ P_{TimeIgnore.DepartAttend.AirlineIgnore} &= 0.1 \times 0.7 \times 0.4 = 0.028,\\ P_{TimeIgnore.DepartIgnore.AirlineAttend} &= 0.1 \times 0.3 \times 0.6 = 0.018,\\ P_{TimeIgnore.DepartIgnore.AirlineIgnore} &= 0.1 \times 0.3 \times 0.4 = 0.012.\end{aligned}$$

Example 3 In the final example, full independence of ANA across attributes is not assumed. Instead, one ANA assignment model handles ANA for flight and departure time, while a second handles ANA for airline only. Thus, ANA is assumed to be independent between airline and the combination of flight and departure time, but need not be between flight and departure time. Additionally, one of the classes, representing the ignoring of flight time and departure time, is not retained. Here, $A = 2$, $|C_1| = 3$, $|C_2| = 2$, and

$$\begin{aligned} C_1 = \{&TimeAttend.DepartAttend, \\ &TimeAttend.DepartIgnore, \\ &TimeIgnore.DepartAttend\} \\ C_2 = \{&AirlineAttend, \\ &AirlineIgnore\} \end{aligned}$$

Dropping one class from C_1 impacts on the cardinality of M, such that $|M| = 6$, and

$$\begin{aligned} M = \{&TimeAttend.DepartAttend.AirlineAttend, \\ &TimeAttend.DepartAttend.AirlineIgnore, \\ &TimeAttend.DepartIgnore.AirlineAttend, \\ &TimeAttend.DepartIgnore.AirlineIgnore, \\ &TimeIgnore.DepartAttend.AirlineAttend, \\ &TimeIgnore.DepartAttend.AirlineIgnore\} \end{aligned}$$

Potential probabilities are presented for each class in each ANA assignment model.

$$\begin{aligned} P_{TimeAttend.DepartAttend} &= 0.65, \\ P_{TimeAttend.DepartIgnore} &= 0.3, \\ P_{TimeIgnore.DepartAttend} &= 0.05, \\ P_{AirlineAttend} &= 0.6, \\ P_{AirlineIgnore} &= 0.4. \end{aligned}$$

Again, the probabilities of each class in the final ANA assignment model are a function of the probabilities in the underlying ANA assignment models.

$$\begin{aligned}
P_{TimeAttend.DepartAttend.AirlineAttend} &= 0.65 \times 0.6 = 0.39, \\
P_{TimeAttend.DepartAttend.AirlineIgnore} &= 0.65 \times 0.4 = 0.26, \\
P_{TimeAttend.DepartIgnore.AirlineAttend} &= 0.3 \times 0.6 \quad = 0.18, \\
P_{TimeAttend.DepartIgnore.AirlineIgnore} &= 0.3 \times 0.4 \quad = 0.12, \\
P_{TimeIgnore.DepartAttend.AirlineAttend} &= 0.05 \times 0.6 = 0.03, \\
P_{TimeIgnore.DepartAttend.AirlineIgnore} &= 0.05 \times 0.4 = 0.02.
\end{aligned}$$

Consider now the choice probabilities conditional on assignment to a class in the final ANA assignment model. While these probabilities can be derived using any form of choice model, the vast majority of latent class ANA models have utilised the MNL model with fixed taste coefficients, which assumes that the unobserved component of utility is independently and identically extreme value type 1 distributed over alternatives and respondents. The formulation here also employs the MNL model, while Section 3.3 utilises the RPL model for the conditional choice probabilities.

The MNL model, without any constraints imposed, will first be defined. Then, the MNL model conditional on assignment to a class in the final ANA assignment model will be introduced, including the specific constraints that will be imposed to reflect ANA. Consider first the total utility of alternative i for respondent n, U_{nit}, which is composed of the representative utility V_{nit}, and the unobserved component of utility, ϵ_{nit}. The representative component is associated with a vector of observed variables, x_{nit}. The utility associated with these variables is estimated with a vector of taste coefficients β, such that the representative utility is $V_{nit} = \beta x_{nit}$. For the MNL model, the probability that alternative i will be chosen is

$$P_{nit} = \frac{e^{\beta x_{nit}}}{\sum_{j=1}^{J} e^{\beta x_{njt}}}. \tag{3.4}$$

The variables that enter into the representative utility contain the K attributes that describe the choice alternatives. Each attribute k may have more than one variable enter into the representative utility, for example if the attribute is dummy or effects coded. The taste coefficients in the β vector represent the sensitivities to the associated variables. For any choice model that is conditioned on a combination of ANA over $K^{\star}$ attributes, some elements of β may be constrained to zero to represent ANA to one or more attributes. Notably, if an attribute is coded such that more than one variable enters into the representative utility, then nonattendance to that attribute is handled by constraining to zero all taste coefficients associated with all variables that represent the attribute (see Scarpa et al., 2009).

Partition the full set of taste parameters β into one or more subsets. First, β_0 is composed of the taste coefficients for the $K - K^\star$ attributes for which ANA is not modelled. This will be an empty set if ANA is modelled for all attributes. Then introduce A subsets, each denoted β_a, which are composed of the taste coefficients associated with the $K_a^\star$ attributes for which ANA is controlled by ANA assignment model a. Each a controls assignment to $|C_a|$ classes, each representing a unique combination of ANA over $K_a^\star$ attributes. Further, each combination will represent a unique pattern of censoring of β_a. For each a, introduce $|C_a|$ sets, each denoted β_{c_a}. The elements of β_{c_a} are either zero, representing ANA, or the taste coefficients drawn from the same position in β_a, representing attendance to the attribute. That is, the taste coefficients that are not censored are constrained to be equal across the C_a sets. Alternatively, unique coefficients could be estimated when censoring does not take place (as with Hensher and Greene, 2010), however, an equality constraint will be imposed in this body of work. The variables to enter into the representative utility, x_{njt}, are similarly partitioned into $A + 1$ subsets. Variables associated with attributes for which ANA is not modelled are in set x_{njt0}, while the variables associated with attributes for which ANA is modelled are partitioned into A subsets x_{njta}.

Conditional on assignment to classes $\{c_1, \dots, c_A\}$ in the each of the A ANA assignment models, the representative utility of alternative j for respondent n on choice occasion t now becomes

$$V_{njt|m} = V_{njt|c_1,\dots,c_A} = \beta_0 x_{njt0} + \sum\nolimits_{a=1}^{A} \beta_{c_a} x_{njta}. \tag{3.5}$$

This censors the taste coefficients associated with the attributes that are ignored in the class of the final ANA assignment model upon which the representative utility is conditioned.

For the MNL model, the probability that respondent n will choose alternative i on choice occasion t, conditional on assignment to classes $\{c_1, \dots, c_A\}$, is

$$P_{nit|m} = P_{nit|c_1,\dots,c_A} = \frac{e^{\beta_0 x_{nit0} + \sum_a^A \beta_{c_a} x_{nita}}}{\sum_{j=1}^{J} e^{\beta_0 x_{njt0} + \sum_a^A \beta_{c_a} x_{njta}}}. \tag{3.6}$$

For panel data, we can specify the probability with respect to a sequence of choices of alternatives over T time periods, $\vec{i} = \{i_1, \dots, i_T\}$. Assuming that the unobserved component of utility is now independently and identically extreme value type 1 distributed over alternatives, respondents, *and* time, the probability of a sequence of choices of alternatives, conditional on assignment to classes $\{c_1, \ldots, c_A\}$, is

$$P_{n\vec{i}|m} = P_{n\vec{i}|c_1,\ldots,c_A} = \prod_{t=1}^{T} \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{c_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{c_a} x_{njta}}} \right]. \tag{3.7}$$

The unconditional probability of a sequence of choices $\vec{i}$ for respondent n is obtained by taking the product of two probabilities: the probability of a combination of ANA, and the probability of the sequence of choices, conditional on assignment to that combination of ANA; then integrating over all analyst specified combinations of ANA. This can be expressed as

$$P_{n\vec{i}} = \sum_{m \in M} P_{nm} P_{n\vec{i}|m} = \sum_{c_1 \in C_1} \cdots \sum_{c_A \in C_A} P_{n\{c_1,\ldots,c_A\}} P_{n\vec{i}|c_1,\ldots,c_A}. \tag{3.8}$$

Substituting in Equations 3.3 and 3.7, Equation 3.8 becomes

$$P_{n\vec{i}} = \sum_{c_1 \in C_1} \cdots \sum_{c_A \in C_A} \prod_{a=1}^{A} \left[\frac{e^{(\gamma_{c_a} + \theta_{nc_a} z_n)}}{\sum_{d \in C_a} e^{(\gamma_d + \theta_{nd} z_n)}} \right] \prod_{t=1}^{T} \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{c_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{c_a} x_{njta}}} \right]. \tag{3.9}$$

Certain specifications of A allow the model to represent the two LC approaches in the literature. If there is only one ANA assignment model, i.e., $A = 1$, then this is a conventional LC model, with specific constraints on the taste coefficients across classes, reflecting ANA. Since this can capture correlation in ANA across all attributes, this extreme will be referred to as the correlated attribute nonattendance (CANA) model. If there is one ANA assignment model for every attribute for which ANA is modelled, i.e., $A = K^\star$, then this is the EAA model from Hole (2011a). This extreme will be referred to as the independent attribute nonattendance (IANA) model. If $1 < A < K^\star$, then this is an ANA model that assumes that independence of ANA holds only between some subsets of the $K^\star$ attributes. This will be referred to as a hybrid ANA model. The hybrid ANA model has not been presented in the literature, and represents one of the contributions of this thesis. In the interest of brevity, the ANA acronyms may be appended by $K^\star$, which represents the number of attributes to which nonattendance is modelled[4]. If $K^\star = 1$, then the single attribute for which ANA is modelled may follow the acronym when referencing the model (e.g., ANA1 fare model).

[4] A similar notation was adopted by Scarpa et al. (2009) for their concept of partial nonattendance (e.g., PNA1).

3.3 The random parameters attribute nonattendance (RPANA) model

To capture preference heterogeneity amongst decision makers that attend to the attributes, we now introduce random parameters, such that the taste coefficients β vary over decision makers with density $f(\beta)$. A distribution is specified for each taste coefficient, and the moments of these distributions are estimated with structural parameters. Most commonly used distributions are described by two moments, however, this thesis employs several distributions for which a single moment is estimated; notably, the Rayleigh distribution, and the constrained triangular and uniform distributions, wherein the spread will be constrained to equal the mean, and so only one moment is actually estimated.

Equation 3.7 now becomes

$$P_{n\vec{i}|m} = P_{n\vec{i}|c_1,\ldots,c_A} = \int \prod_{t=1}^{T} \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{c_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{c_a} x_{njta}}} \right] f(\beta) d\beta. \tag{3.10}$$

Substituting Equation 3.10 into Equation 3.8, we obtain an unconditional probability of a sequence of choices $\vec{i}$ for respondent n of

$$P_{n\vec{i}} = \sum_{c_1 \in C_1} \cdots \sum_{c_A \in C_A} \prod_{a=1}^{A} \left[\frac{e^{(\gamma_{c_a} + \theta_{nc_a} z_n)}}{\sum_d^{C_a} e^{(\gamma_d + \theta_{nd} z_n)}} \right] \int \prod_{t=1}^{T} \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{c_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{c_a} x_{njta}}} \right] f(\beta) d\beta. \tag{3.11}$$

This choice probability underpins the RPANA model.

Like the ANA model, the RPANA model also has two extremes of correlation of ANA. The random parameters independent attribute nonattendance (RPIANA) model is a RPANA model in which $A = K^\star$, and which relies on the assumption that ANA is independent across attributes. The random parameters correlated attribute nonattendance (RPCANA) model is a RPANA model in which $A = 1$, and which can capture correlation in ANA across all attributes. If $1 < A < K^\star$, then this is a hybrid RPANA model, which assumes that independence of ANA holds only between some subsets of the $K^\star$ attributes.

3.3.1 Nesting structure of the random parameters attribute nonattendance model

The RPL, ANA and MNL models all nest within the RPANA model. Figure 3.1 depicts the nesting structure. There are two conditions which allow the RPANA model to collapse to the RPL, ANA and MNL models.

1. The probability of full attendance approaches one. Specifically, $P_{n\text{FullAttendance}} \rightarrow 1, \forall n$.

2. All RP distributions become degenerate, and thus point estimates.

Figure 3.1 maps these conditions to the specific model transitions. The RPANA model under full attendance becomes the RPL model; the RPANA model with no preference heterogeneity becomes the ANA model; and the RPANA model under full attendance and with no preference heterogeneity becomes the MNL model. The nesting structure is important, as likelihood ratio tests can only be performed if two models nest. Practical issues with estimating the RPANA model will now be discussed.

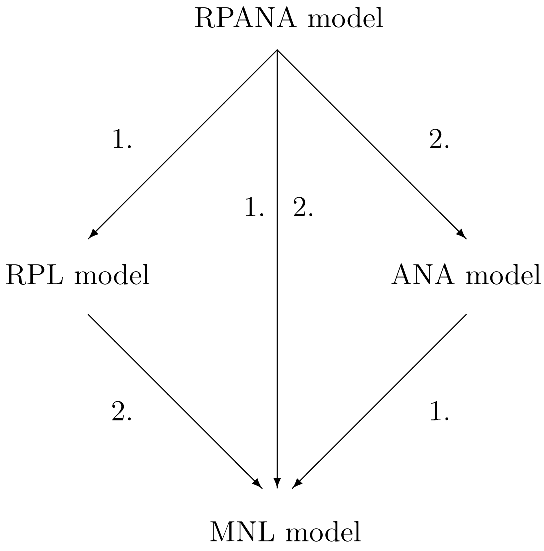

Figure 3.1: Nesting structure of the RPANA, RPL, ANA and MNL models

3.3.2 Identification of the random parameters attribute nonattendance model

The performance of the RPANA model will be examined extensively with simulated data in Chapter 5, and with empirical data in Chapter 6. However, across multiple datasets, it is found that the RPANA model is potentially susceptible to a number of identification problems. Consequently, this section will detail the problems, offer solutions, and note the limitations of the ways in which the RPANA model can be applied. This section can then be drawn upon in subsequent chapters, so that these chapters focus on the specific consequences for the datasets contained therein.

Two types of estimation problems are encountered. The less problematic of these are cases whereby the model converges on a local maxima, which is plausible in the context of such a highly nonlinear model. This was found to be more common when attendance to multiple attributes was being modelled, and typically manifested itself as nonattendance rates tending to zero. In most cases, such problems were overcome by first estimating nonattendance to one attribute at a time, then using the recovered parameter values as start values for the RPANA model that models attendance to multiple attributes. Therefore, caution must be warranted before concluding that ANA rates for an attribute are indeed zero.

A more fundamental problem is concerned with the choice of distribution, and what is believed to be a fundamental incompatibility between the RPANA model and parameter distributions that can span both the positive and negative domain. Across a range of real and simulated datasets, any attempt to include distributions that may span both the positive and negative domain led to a multitude of estimation problems, including flat log likelihoods, very large standard errors, and singular covariance matrices. Problematic distributions include the normal, which is unbounded and by definition will always have support over both the positive and negative domain; the triangular, which is bounded but can freely span zero; and the uniform, also bounded but free to span zero.

Interestingly, the censored normal also exhibits the same problems. With its point mass at zero, the censored normal can already capture ANA. The motivation for the RPANA model over simply using the censored normal is that the latter is likely more prone to confounding ANA with preference heterogeneity, since ANA is captured through the same parameters that capture preference heterogeneity. This will be demonstrated using simulations in Section 4.3.5. The unbounded nature of the underlying normal distribution suggests that the ANA rate implied by the censored normal distribution is always greater than zero. If it is very close to zero, through some appropriate combination of μ and σ, then the RPANA model could capture the vast majority of ANA, and the censored normal distribution would primarily capture the continuous component of utility. What appears to be happening in practice is that the potential to capture ANA through both the ANA parameter, and the censoring of the normal distribution, leads to an identification problem whereby some arbitrary combination of the two sources of ANA can approximate the 'true' ANA. This in turn leads to the problems with estimation.

The same phenomenon may be occurring with the normal, triangular and uniform distributions. Now, however, a certain proportion of coefficients close to zero, including those of implausible sign, is approximating ANA. This in turn leads to an identification problem, with the ANA parameter and the continuous distribution's support near zero both 'competing' for the share of attribute nonattenders. By limiting the support of the continuous distribution near zero through the application of a distribution that is bounded on one side at zero, this identification problem can potentially be overcome. Distributions that appear to work well across a variety of datasets include the constrained triangular, lognormal, and Rayleigh.

However, the use of a zero bounded distribution appears to be a necessary but not sufficient condition. Problems are encountered with the dataset used in Chapter 6 with the constrained uniform distribution, in which the spread is constrained to be equal to the mean. This results in an equal share of coefficients over a domain spanning between zero and two times the mean. It may be that by not tapering towards zero, the continuous distribution has enough support near zero to suitably approximate ANA, leading to an identification problem. The consequence of this is that care must be taken when choosing distributions, and the specifics of any empirical application may have an impact on what can be identified. To some extent, this also calls into question the confidence the analyst can place on an inferred ANA rate. Indeed, it may not be possible to completely unentangle ANA and low attribute sensitivity.

Use of a zero bounded distribution poses problems for attributes for which taste coefficients of both sign are plausible. For example, Train (1998) estimated a normally distributed random parameter with an insignificant mean and significant standard deviation, and found that some anglers preferred fishing sites with campgrounds, while others preferred sites without. In this example, it might be that, additionally, some fishers are actually indifferent to the campground attribute, and will pay no attention to it. To some extent the normal distribution will represent such fishers through the distribution's support near zero. Indeed, the mode of the distribution is close to zero. However, the normal distribution might do a poor job of approximating preference heterogeneity for such an attribute if there is a high ANA rate and strong preferences for and against the attribute. Capturing the strong sensitivities might compromise capturing ANA. As with the censored normal, the potential problem stems from a distribution with two moments being utilised to try and capture three effects: measures of both central tendency and dispersion, *and* indifference to the attribute. No solution for applying the RPANA model to such attributes is offered.

A solution is offered for a related problem, however. In many choice model specifications, categorical or ordinal attributes will enter into the utility function through dummy or effects coded variables. To take an example from the empirical application of Chapter 6, departure time for a scheduled flight could be dummy coded such that taste coefficients are estimated for specific departure times, or ranges thereof. One attribute level or range is normalised to facilitate estimation, and the estimated coefficients are relative to the normalised value.

However, one attribute level will not necessarily be preferred to another by all decision makers. For example, a 10am departure time may be preferred to 6am by most decision makers, but some may prefer the early departure, perhaps because it allows for a longer day at the destination. Random parameters allow this to be captured by estimating distributions for each attribute level that overlap in their domain. However, if dummy coding is employed, and the base level is normalised to zero, then unless this attribute level is universally preferred or not preferred over another attribute level, the RP distribution for that second attribute level will span zero. Consequently, an identification problem may result if nonattendance to the attribute is estimated with the RPANA model. However, this problem appears to be overcome by jointly censoring each of the parameters associated with the dummy coded attribute. That is, if a dummy coded attribute is ignored, then this is best represented by constraining all associated parameters to zero, as the entire attribute is ignored, not just particular levels. The optimism that the model can be identified stems from the likelihood that the ANA condition is harder to approximate with a number of independently varying random parameters that are each associated with an attribute level.

In a number of empirical applications, it is found that the introduction of ANA jointly across all parameters associated with a dummy coded attribute leads to an improvement in model fit. However, model estimation is not stable, with singular covariance matrices commonly occurring, suggesting that an identification problem may remain. One potential source of the problem is the normalisation of the dummy parameters and ASCs, where one coefficient is fixed to zero. An alternative normalisation can be achieved with effects coding. An attribute k with L_k levels is coded into $L_k - 1$ variables. A utility coefficient, β_{kl}, is estimated for each of these variables. The base level of utility is not zero, as with dummy coding, but $\sum_l^{L_k-1} -\beta_{kl}$. Crucially, with effects coding employed, no estimation problems are encountered. Effects coding the ASCs is unusual, however Train (2009) notes that the ASCs need not be normalised to zero, and that doing so is merely easier. With the RPANA model, we have sufficient motivation to deviate from convention. Gilbride et al. (2006) noted that dummy coding cannot be used, in their Bayesian stochastic attribute selection model. Instead they implemented orthogonal coding, which like effects coding does not utilise base levels of zero. It appears that there may be confounding between the inferred ANA and the zero base levels of dummy coded attributes. Alternative coding schemes are a ready solution.

A large number of draws were used in all models estimated in this thesis. This was motivated by a concern that too few draws might mask identification problems in model estimation (Walker, 2002). The concern was heightened by some observed problems with the

RPANA model with the constrained uniform distribution. Some inconsistency in convergence was observed for this distribution with a small number of draws. Five thousand draws was found to be a suitable number, and applied to all random parameter models estimated in this thesis, for consistency. This is, however, a relatively large number compared to that typically applied in the literature (see for example Hensher and Greene, 2003).

3.4 Summary of attribute nonattendance models in the literature

Table 3.1 summarises some of the key papers in the literature that have utilised some form of the latent class based ANA model. They are categorised on the form of the model, as defined in this chapter (e.g., CANA); the number of attributes, $K^{\star}$, for which ANA was modelled; whether random parameters were employed; whether the coefficients were constrained to be equal in each class that they were not set to zero; and whether covariates were introduced to vary the ANA probabilities across respondents. Hess et al. (2011) limited their application by assuming full independence of ANA across the attributes, and sidestepped identification issues by only using a lognormal distribution, on attributes for which more or less is clearly better. Unlike in this thesis, Hensher et al. (2012b) did not find improvements in model fit with the introduction of random parameters to the ANA model. The entry for this thesis relates to the empirical application contained in Chapter 6. The contributions over the proceeding two papers that also employ RP distributions include, but are not limited to, estimation of a hybrid[5] RPANA model, employment of multiple distributions, exploration of identification issues, and the introduction of covariates into the ANA assignment models.

Table 3.1: Summary of ANA models in the literature

Paper	**Model**	**K***	**Random parameters**	**Equality constraint**	**ANA covariates**
Hess and Rose (2007)	ANA	1	No	N/A	Sociodem.
Scarpa et al. (2009)	CANA	5	No	Yes	-
Hensher and Greene (2010)	CANA	4	No	No	-
Hole (2011a)	IANA	5	No	Yes	-
Hess et al. (2011)	RPIANA	2/5/6	Yes	Yes	-
Hensher et al. (2012b)	RPCANA	3	Yes	No	-
This thesis	RPANA hybrid	4	Yes	Yes	Stated ANA

[5]Specifically, $1 < A < K^{\star}$.

Chapter 4

Quantifying the impacts of attribute nonattendance with simulations

4.1 Introduction

A risk of not adequately handling ANA is that key model outputs such as taste parameters and WTP measures may be biased (see Section 2.2.3 for a review of these findings). Further, Hensher (2007) has shown that handling ANA may decrease the incidence in RPL models of coefficients of behaviourally implausible sign (also see Section 2.2.3 for details). However, most findings in the literature are constrained by an empirical setting, which only allows a limited number of data points to be observed, and which raises difficulties in measuring bias, as the truth cannot be known with certainty. There remains a lack of precise knowledge about the nature of these biases and sign violations, in terms of their severity, the direct influence of the ANA rate, and the interaction of ANA with other phenomena such as taste heterogeneity. An understanding of these biases and sign violations is important, because it provides some sense of just how misspecified a model may be if ANA is not handled satisfactorily.

Empirical studies are of limited value in obtaining a deep understanding of the biasing influences of ANA. A measure of bias in this context requires a comparison of the true state with the state that is naïve to ANA. However, in any empirical context, it is impossible to know with certainty what the true state is. The model outputs after ANA has been handled cannot be relied upon as being the true state, as the method of identifying ANA might itself have introduced bias. The problems with stated and analytically derived ANA will be considered in turn.

Statements made by the respondent about whether they ignored an attribute (e.g., Hensher et al., 2005; Campbell et al., 2008) will only be an approximation of the truth, and may contain not just error at the level of the individual, but, potentially, bias across a sample. Indeed, numerous studies have claimed that those who state that they do not attend to an attribute merely have a lower sensitivity to that attribute (Hess and Rose, 2007; Carlsson et al., 2010), evidenced by the estimation of significant parameters for respondents that claim they ignored an attribute. An alternative explanation for the same outcome is that, of those who claim to ignore an attribute, some truly do ignore it, while others have some degree of sensitivity to it. Irrespective of the interpretation, it has been shown that stated ANA will likely contain some degree of error.

Analytical methods could be used to identify ANA, and so compare the results with models that are naïve to ANA. Examples include the use of conditional parameter estimates (Hess and Hensher, 2010), the CANA model (Hess and Rose, 2007; Scarpa et al., 2009), and the IANA model (Hole, 2011a). However, the analyst cannot be certain that these methods are correctly capturing ANA. Overall, the use of empirical data for quantifying the impact of ANA on model outputs is highly problematic.

An appealing alternative is to test for the biasing impact of ANA with simulated datasets. By artificially generating choices across a simulated sample of respondents, control can be exerted over the true patterns of nonattendance and preference heterogeneity. These variables can be systematically varied, allowing a wide range of potential values to be tested, and patterns in the biases to be more readily discerned. Furthermore, since the true values are known, not latent, precise measures of bias can be calculated[1]. In this way, the extent of potential biases arising from ANA can be catalogued and measured.

Several studies have used simulations to investigate the biasing impact of ANA. Hoyos et al. (2010) utilised datasets generated with fixed coefficients, and found that the estimated sensitivities were biased downwards by the same percentage as the ANA rate. Mariel et al. (2011) estimated RPL models on datasets generated with continuous distributions of coefficients, and varying degrees of ANA. The means of the RP distributions were biased downwards and the standard deviations upwards. However, the true standard deviations were not varied, nor were the shapes of the distributions. Rose et al. (2012b) examined lexicographic behaviour, which is a special case of ANA whereby all attributes except for one are ignored. As the incidence rate of lexicographic choice with respect to an attribute increased, so too did the magnitude of both the mean and standard deviation of the parameter associated with

[1]Subject to some degree of simulation noise, which can be mitigated by generating a large number of simulated datasets.

that attribute.

This chapter will explore the bias in the mean and dispersion[2] of RP distributions, as well as the incidence of sign violation, as a number of dimensions are systematically varied in the simulated datasets. As with previous studies, the ANA rate is systematically varied, for it is the key influence of interest. Additionally, the dispersion of the distribution is varied, with three specifications, ranging from a narrow to a wide range of coefficients. A number of distribution shapes are tested, to see if the shape of the distribution has some impact on the biasing influence of ANA. By varying a number of dimensions appropriately, interactions between these dimensions can also be tested for. Varying the shape and dispersion of the distribution is important, as it impacts upon the amount of mass near zero. If there is confounding between ANA and low taste sensitivities, as suggested by Campbell et al. (2012) and Hess et al. (2011), then these simulations are well specified to allow such confounding to be detected and investigated.

The motivation thus far has been the quantification of the potential biases induced when using ubiquitous methodologies such as the RPL model, that are naïve and potentially not well suited to a mass of respondents who are indifferent to an attribute. Numerous analytical methods have been proposed for measuring and accommodating ANA. However, in most cases, these methods have only been applied to empirical datasets (the notable exception being Mariel et al. (2011), discussed below). An important question is whether these methods accurately measure and accommodate ANA, or introduce their own biases. Comparison with stated ANA, if available, provides some degree of face validity, however, the accuracy of stated ANA cannot be guaranteed, as already discussed. Simulated datasets allow the accuracy of the analytical methods to be tested accurately, as the retrieved model estimates can be compared to their true values. This chapter will investigate two techniques for capturing ANA analytically: the RPL model with a censored normal distribution, and the IANA model. Similar findings were obtained with the IANA and CANA models; only results from the former will be reported. Hoyos et al. (2010) did investigate the CANA model, but simulated their datasets with fixed taste coefficients. This chapter will use simulated datasets generated with distributions of taste coefficients, so that the impact of taste heterogeneity on the accuracy of the IANA model can be tested. The datasets generated will also serve as a test bed for the model introduced in this body of work, the RPANA model, the results of which will be presented in Chapter 5.

The performance of the conditional parameter estimate technique for handling ANA will

[2]The standard deviation or the spread, depending on the distribution employed.

not be examined in the chapter, as Mariel et al. (2011) have investigated this approach using simulated datasets. They found that the accuracy with which an attribute is estimated as nonattended is dependant on the threshold value of the coefficient of variation that the conditional parameter estimate approach requires, and that the best performing threshold value varies as the true ANA rate varies. However, the true ANA rate is latent in empirical applications. Consequently, the analyst cannot be certain of what threshold value to set, and so also of whether the estimated ANA rate is accurate.

A caveat must be issued on the simulations performed in this chapter. The simulations are highly controlled, with choices following very precise rules. Empirical datasets might exhibit less predictable choice patterns, including, but not limited to: random choice, fatigue effects, extreme sensitivities to attributes, noncompensatory behaviour, and preference distributions that differ from those examined in this chapter. Any of these unpredictable choice patterns might distort the biases measured in the controlled environment of the simulations, as well as the assessments of the accuracy of the existing analytical methodologies for handling ANA. This does not undermine the proposed research agenda, as the simulations nonetheless provide the analyst with a great deal of insight.

4.2 Methodology

In this section, the simulation methodology will be outlined. First, the methods for constructing the simulated datasets will be presented. This spans the utility specifications applied, the RP distributions tested, and, crucially, the experimental design applied to ensure variability in ANA and other suspected biasing influences. Next, the specifications of the models estimated on the simulated datasets are detailed. Finally, the measures of bias are introduced, together with methods for analysing the biasing influences.

4.2.1 Data generation

The simulated choice task contains two alternatives, each described by three attributes. Each of the two alternatives has the following utility specification:

$$V_j = \beta_1 attribute_{j1} + \beta_2 attribute_{j2} + \beta_3 attribute_{j3}, \qquad j = 1, 2 \tag{4.1}$$

where $attribute_{j1} \in \{1, 3, 5, 7\}$, $attribute_{j2} \in \{15, 20, 25, 30\}$ and $attribute_{j3} \in \{0, 1\}$. All three parameters are specified as random parameters, where the distribution applied is consistent across all parameters within a dataset, but may vary across datasets. Normal, triangular, uniform, and lognormal distributions are applied, where each distribution is specified with two parameters: a mean μ, and a measure of dispersion σ. The measure of dispersion represents a spread for the triangular and uniform distributions, and a standard deviation for the normal and censored normal distributions. For the lognormal distribution, μ and σ represent the mean and standard deviation of the underlying normal distribution, i.e., $ln\beta_k \sim N(\mu, \sigma)$.

Since it is hypothesised that the magnitude of the true random parameter dispersion may interact with the rate of nonattendance in introducing bias to the estimated distribution, the normalised measure of dispersion is systematically varied, both across datasets, and across parameters within each dataset. Three levels of dispersion are tested for each parameter, and are detailed in Table 4.1. Each combination of parameter and dispersion level will be referred to as a parameter specification in this chapter. For the the normal, triangular and uniform distributions, σ takes on levels $\frac{1}{3}\mu$, $\frac{2}{3}\mu$, and μ. These levels loosely represent conditions of low, medium and high dispersion, and will be referred to as dispersion specifications in this chapter. For the triangular and uniform distributions, in the high dispersion specification, where $\mu = \sigma$, preference heterogeneity is bounded by zero, and so in none of the specifications does the true distribution have coefficients with differing sign to the mean of the distribution, referred to herein as sign violation[3]. In particular, the high dispersion specification presents a distribution that can be parsimoniously estimated with a constrained distribution, whereby a single parameter is estimated that represents both the mean and the spread (see Hensher and Greene, 2003). The normal distribution has an unbounded domain, and so will always generate coefficients of both signs with a nonzero probability[4]. Table 4.1 contains a column that indicates the percentage of such coefficients, which ranges from 0.1 percent for the low dispersion specification to 15.9 percent for the high dispersion specification. This complicates the analysis of the impact of ANA on the percentage of sign violation, since the true percentage must also be accounted for.

For the lognormal distribution, the parameters of the underlying normal distribution, μ and σ, are set with consideration of the statistical measures of the lognormal distribution. Four such measures are presented in Table 4.1: the mean ($e^{\mu+\sigma^2/2}$), the median (e^{μ}), the mode ($e^{\mu-\sigma^2}$), and the standard deviation ($\sqrt{(e^{\sigma^2}-1)e^{2\mu+\sigma^2}}$). For each parameter, the mean of

[3]Note that in many empirical contexts, both positive and negative coefficients may be behaviourally valid, with some respondents preferring more of an attribute, and others preferring less. In the simulated choices with the uniform, triangular, lognormal and censored normal distributions, it shall be assumed that for each parameter, only one sign is behaviourally valid.

[4]For the normal distribution, it shall be assumed that some percentage of sign violation is behaviourally plausible.

Table 4.1: Parameter distributions that enter the utility specification

Parameter 1						
Normal, triangular, uniform	μ	σ	β_1>0 (%)			
	-0.3	0.1	0.1			
	-0.3	0.2	6.7			
	-0.3	0.3	15.9			
Lognormal	μ	σ	mean	median	mode	std dev.
	-1.265	0.349	0.3	0.282	0.25	0.193
	-1.339	0.520	0.3	0.262	0.20	0.208
	-1.435	0.680	0.3	0.238	0.15	0.229
Censored normal	μ	σ	truncated mean	truncated std dev.		
	-0.300	0.1	0.3	0.099		
	-0.263	0.2	0.3	0.170		
	-0.144	0.3	0.3	0.208		
Parameter 2						
Normal, triangular, uniform	μ	σ	β_2>0 (%)			
	-0.06	0.02	0.1			
	-0.06	0.04	6.7			
	-0.06	0.06	15.9			
Lognormal	μ	σ	mean	median	mode	std dev.
	-2.842	0.241	0.06	0.058	0.055	0.037
	-2.909	0.438	0.06	0.055	0.045	0.040
	-2.993	0.599	0.06	0.050	0.035	0.044
Censored normal	μ	σ	truncated mean	truncated std dev.		
	-0.060	0.02	-0.06	0.020		
	-0.053	0.04	-0.06	0.034		
	-0.029	0.06	-0.06	0.042		
Parameter 3						
Normal, triangular, uniform	μ	σ	β_3<0 (%)			
	0.9	0.3	0.1			
	0.9	0.6	6.7			
	0.9	0.9	15.9			
Lognormal	μ	σ	mean	median	mode	std dev.
	-0.231	0.680	1	0.794	0.5	0.764
	-0.350	0.837	1	0.705	0.35	0.861
	-0.536	1.036	1	0.585	0.2	1.037
Censored normal	μ	σ	truncated mean	truncated std dev.		
	0.899	0.3	0.9	0.298		
	0.789	0.6	0.9	0.510		
	0.433	0.9	0.9	0.624		

the lognormal distribution remains the same, with the standard deviation varying across the three specifications. It should be noted that while holding the mean constant, an increase in variance of the lognormal distribution will lead to a decrease in median and mode of the distribution. The changing mode is clearly evident in the plots of the lognormal distributions presented in Figure 4.1[5]. As the variance increases for each parameter, the modes decrease, the tails of the distributions become fatter, and the distribution becomes increasingly skewed to the right.

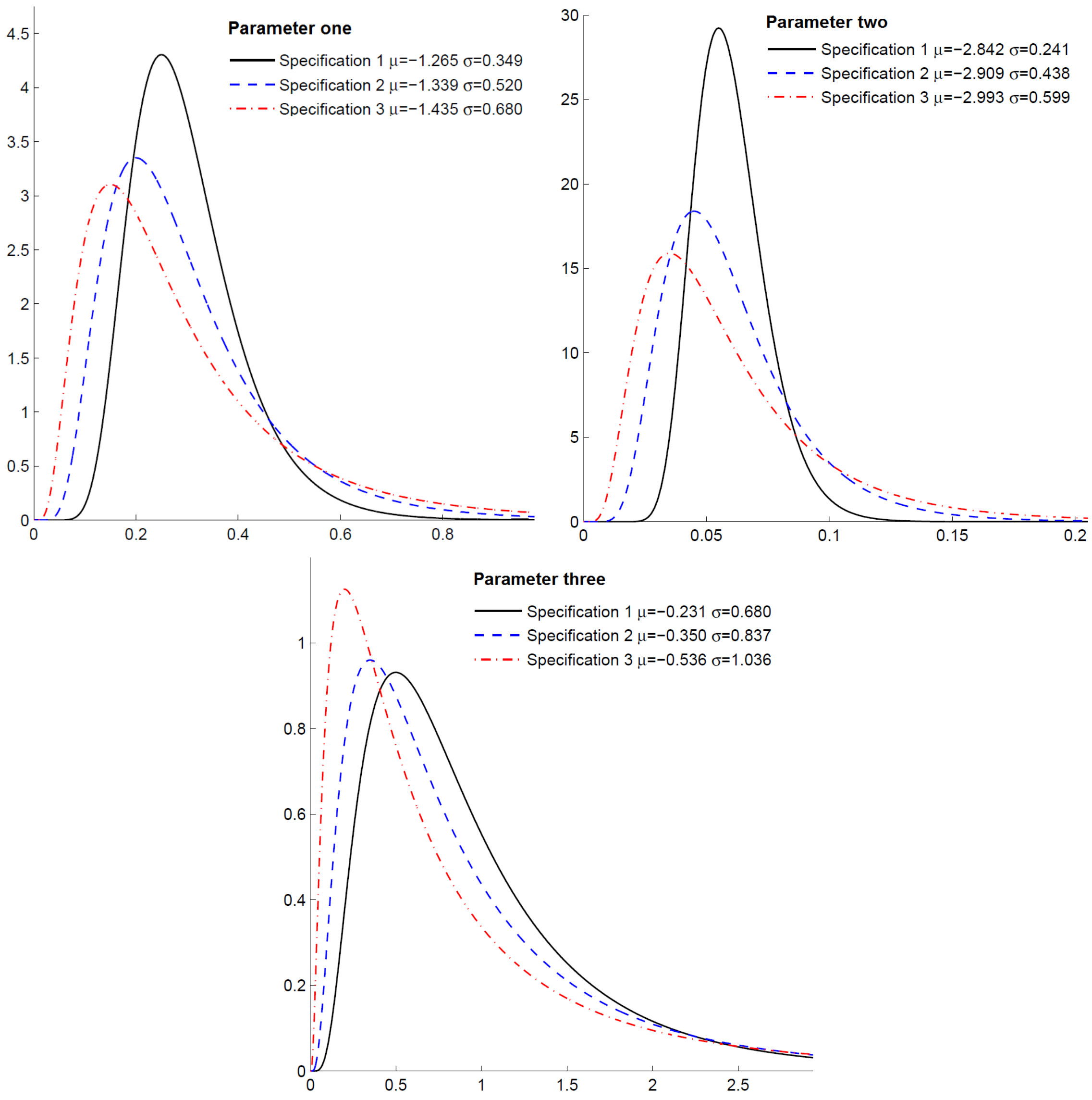

Figure 4.1: True lognormal distributions

[5]The signs of the coefficients of parameters one and two have been reversed in the plots, as the domain of the lognormal distribution is strictly positive. The attributes associated with these parameters have their sign reversed prior to estimation. The estimated distribution can then have the sign of its coefficients reversed for interpretation.

As with the lognormal distribution above, the censored normal distribution will have moments that differ from the underlying normal distribution. A censored normal distribution contains both a continuous component (representing preference heterogeneity in the RPL model) and a discrete component (representing indifference or nonattendance to an attribute in the RPL model, when the censoring point is zero). However, when constructing the datasets with censored normal RP distributions, and when analysing the ability of such a distribution to accurately capture ANA, of interest are the moments of the continuous component of the distribution, which describe the preference heterogeneity of those that attend to the attribute. The point mass at zero representing ANA can be analysed separately. The truncated normal distribution is convenient, as the mean and variance of this distribution will represent the mean and variance of the continuous component of the censored normal distribution, if the moments of the underlying normal distribution are the same. Greene (1997) details the calculation of the mean and variance of the truncated normal distribution of coefficients β, for truncation point b, with an underlying normal distribution with mean μ and variance σ^2:

$$
\begin{aligned}
Mean[\beta|\text{truncation}] &= \mu + \sigma\lambda(\alpha), \\
Variance[\beta|\text{truncation}] &= \sigma^2[1 - \delta(\alpha)], \\
&\text{where} \\
&\quad \alpha = (b - \mu)/\sigma, \\
&\quad \lambda(\alpha) = \phi(\alpha)/[1 - \Phi(\alpha)] \qquad \text{if truncation is } \beta > b, \\
&\quad \lambda(\alpha) = -\phi(\alpha)/\Phi(\alpha) \qquad \text{if truncation is } \beta < b, \\
&\quad \delta = \lambda(\alpha)[\lambda(\alpha) - \alpha], \\
&\quad \phi(.) \quad \text{is the standard normal pdf, and} \\
&\quad \Phi(.) \quad \text{is the standard normal cdf.}
\end{aligned}
\tag{4.2}
$$

Over the three specifications of each of the three parameters, the mean of the truncated normal is kept constant in the simulations. Again there is a low, medium and high dispersion specification, with the truncated standard deviation diminished relative to the standard deviation of the underlying normal distribution, σ. The specifications are detailed in Table 4.1.

Central to the systematic testing of the impact of ANA on choice model results is the variation in ANA rates across the simulation runs performed. Four rates were selected: zero, 15, 30, and 45 percent. Zero percent is important, as it facilitates testing of the ability of the choice model to recover the true parameter estimates in the absence of ANA. An upper bound of 45 percent was selected, as it was deemed that a higher rate might lead to problems recovering preference heterogeneity, especially in the RPANA models estimated on the simulated datasets in Chapter 5. These are exogenous assignments of nonattendance for the relevant percentage of respondents.

A d-efficient experimental design (Huber and Zwerina, 1996; Rose et al., 2008) was generated to provide the specific attribute levels faced by each simulated respondent in each choice task. The means of the normal (and hence triangular and uniform) distributions, as presented in Table 4.1, were used as fixed parameter priors. The experimental design was generated with 20 choice tasks, with each simulated respondent completing all 20 tasks, and the same choice tasks being completed by all respondents. With 400 simulated respondents, each dataset contained 8000 observations. It is unclear what impact the specific design strategy will have on the outcomes of the simulations. For example, the use of different parameter priors, or the generation of orthogonal designs, might have an impact on the results. With no firm hypothesis, and an already burdensome number of dimensions varied for their impact on ANA, it was decided to use a single, d-efficient design, which represents current best practice.

For any given dataset, decisions needed to be made on a number of dimensions. These included the distribution of the random parameters, the selection of structural parameter values controlling the RP distribution associated with each attribute k (μ_k and σ_k), and the ANA rate for each attribute (ANA_k). A combination of these variables will be referred to as a treatment, where the systematic generation of the treatments will be discussed shortly. However, for any given treatment, the choice alternative chosen needed to be determined for each choice task in the dataset. First, specific coefficients β_{nk} needed to be drawn for each random parameter, for each simulated respondent n, from the appropriate distribution. It was assumed that sensitivity to the attributes would remain constant for each respondent across all 20 choice tasks. That is, the panel nature of the choice tasks was accounted for at the data generation stage. Whilst some studies have suggested that this assumption may not always hold empirically (Hess and Rose, 2009), it is a reasonable assumption to make in the context of this study. Second, the ANA rate assigned to each parameter, ANA_k, was used to randomly censor to zero the coefficients β_{nk} of that percentage of the 400 respondents. Third, a draw ϵ_{nit} (where t is the choice occasion and i the choice alternative) was taken from a gumbel distribution, to represent unobserved influences on the choice. The use of a gumbel distribution is consistent with the logit models employed in this chapter. Finally, the utility for each choice alternative was calculated as a summation of the three parameter coefficients and the gumbel draw, and the alternative in each choice set with the highest utility was assigned as the chosen alternative.

With the candidate RP distributions, parameter values and ANA rates defined, the next step was to define how these variables were to be systematically varied across multiple datasets. An orthogonal design was generated, such that no correlations existed between the the ANA rate and the dispersion specification (low, medium and high), across all three parameters. Such an approach sought to isolate the impact of varying each of the variables. Note that the experimental design only varied the ANA rate and dispersion specification. The same experimental design was applied to each of the five distributions, with σ assuming the role of either standard deviation (for the normal and censored normal distributions) or spread (for the triangular and uniform distributions), and μ and σ assuming the appropriate values outlined in Table 4.1 for the low, medium and high dispersion specifications for the log-normal and censored normal distributions. This decision to use the same experimental design for each distribution is justified because the impact of ANA and preference heterogeneity on parameter estimates will be analysed for each distribution independently. A full discussion of the strategy for modelling these impacts will be presented in Section 4.2.3. The orthogonal experimental design is presented in Table 4.2 for the normal, uniform and triangular distributions, and Table 4.3 for the lognormal distribution. The experimental design contains 36 treatments. Of particular note are the first three treatments, which each have full attendance for all three attributes. These represent treatments untainted by ANA, and as such we would expect the naïve RPL choice models to recover the parameters with a reasonable degree of accuracy.

This in turn raises the issue of noise in the data generation process. Numerous aspects of the dataset generation process were random, including the draws of the random parameter coefficients from the densities, the draws from the gumbel distribution, and the selection of which respondents ignore each attribute. This randomness introduced noise into the simulation exercise. For example, in any one dataset, the draws from a density may provide a poor approximation to that density. To integrate out this noise, 100 datasets were generated for each combination of treatment and RP distribution, with all parameter values then averaged across these 100 datasets before being analysed. With five distributions, 36 treatments, and 100 datasets per distribution/treatment combination, 18400 data sets were generated. An even greater number of models were estimated, with multiple models estimated per dataset, as outlined in the next section.

Table 4.2: Treatments for normal, uniform and triangular distributions

	Parameter 1				Parameter 2				Parameter 3			
Tr.	μ	σ	**>0** (%)	**ANA** (%)	μ	σ	**>0** (%)	**ANA** (%)	μ	σ	**<0** (%)	**ANA** (%)
1	-0.3	0.1	0.1	0	-0.06	0.02	0.1	0	0.9	0.6	6.68	0
2	-0.3	0.2	6.7	0	-0.06	0.04	6.7	0	0.9	0.9	15.87	0
3	-0.3	0.3	15.9	0	-0.06	0.06	15.9	0	0.9	0.3	0.13	0
4	-0.3	0.2	6.7	30	-0.06	0.06	15.9	45	0.9	0.9	15.87	30
5	-0.3	0.3	15.9	30	-0.06	0.02	0.1	45	0.9	0.3	0.13	30
6	-0.3	0.1	0.1	30	-0.06	0.04	6.7	45	0.9	0.6	6.68	30
7	-0.3	0.3	15.9	15	-0.06	0.06	15.9	15	0.9	0.3	0.13	45
8	-0.3	0.1	0.1	15	-0.06	0.02	0.1	15	0.9	0.6	6.68	45
9	-0.3	0.2	6.7	15	-0.06	0.04	6.7	15	0.9	0.9	15.87	45
10	-0.3	0.1	0.1	30	-0.06	0.04	6.7	0	0.9	0.6	6.68	45
11	-0.3	0.2	6.7	30	-0.06	0.06	15.9	0	0.9	0.9	15.87	45
12	-0.3	0.3	15.9	30	-0.06	0.02	0.1	0	0.9	0.3	0.13	45
13	-0.3	0.3	15.9	15	-0.06	0.02	0.1	30	0.9	0.9	15.87	15
14	-0.3	0.1	0.1	15	-0.06	0.04	6.7	30	0.9	0.3	0.13	15
15	-0.3	0.2	6.7	15	-0.06	0.06	15.9	30	0.9	0.6	6.68	15
16	-0.3	0.2	6.7	45	-0.06	0.06	15.9	45	0.9	0.6	6.68	0
17	-0.3	0.3	15.9	45	-0.06	0.02	0.1	45	0.9	0.9	15.87	0
18	-0.3	0.1	0.1	45	-0.06	0.04	6.7	45	0.9	0.3	0.13	0
19	-0.3	0.3	15.9	45	-0.06	0.04	6.7	0	0.9	0.6	6.68	0
20	-0.3	0.1	0.1	45	-0.06	0.06	15.9	0	0.9	0.9	15.87	0
21	-0.3	0.2	6.7	45	-0.06	0.02	0.1	0	0.9	0.3	0.13	0
22	-0.3	0.3	15.9	45	-0.06	0.04	6.7	30	0.9	0.9	15.87	45
23	-0.3	0.1	0.1	45	-0.06	0.06	15.9	30	0.9	0.3	0.13	45
24	-0.3	0.2	6.7	45	-0.06	0.02	0.1	30	0.9	0.6	6.68	45
25	-0.3	0.1	0.1	0	-0.06	0.06	15.9	45	0.9	0.3	0.13	15
26	-0.3	0.2	6.7	0	-0.06	0.02	0.1	45	0.9	0.6	6.68	15
27	-0.3	0.3	15.9	0	-0.06	0.04	6.7	45	0.9	0.9	15.87	15
28	-0.3	0.2	6.7	0	-0.06	0.04	6.7	30	0.9	0.3	0.13	30
29	-0.3	0.3	15.9	0	-0.06	0.06	15.9	30	0.9	0.6	6.68	30
30	-0.3	0.1	0.1	0	-0.06	0.02	0.1	30	0.9	0.9	15.87	30
31	-0.3	0.2	6.7	30	-0.06	0.02	0.1	15	0.9	0.3	0.13	15
32	-0.3	0.3	15.9	30	-0.06	0.04	6.7	15	0.9	0.6	6.68	15
33	-0.3	0.1	0.1	30	-0.06	0.06	15.9	15	0.9	0.9	15.87	15
34	-0.3	0.1	0.1	15	-0.06	0.02	0.1	15	0.9	0.9	15.87	30
35	-0.3	0.2	6.7	15	-0.06	0.04	6.7	15	0.9	0.3	0.13	30
36	-0.3	0.3	15.9	15	-0.06	0.06	15.9	15	0.9	0.6	6.68	30

Table 4.3: Treatments for lognormal distribution

	Parameter 1					Parameter 2					Parameter 3				
Tr.	**mean**	**median**	**mode**	**s.d.**	**ANA (%)**	**mean**	**median**	**mode**	**s.d.**	**ANA (%)**	**mean**	**median**	**mode**	**s.d.**	**ANA (%)**
1	0.3	0.282	0.25	0.193	0	0.06	0.058	0.055	0.037	0	1	0.705	0.35	0.861	0
2	0.3	0.262	0.20	0.208	0	0.06	0.055	0.045	0.040	0	1	0.585	0.20	1.037	0
3	0.3	0.238	0.15	0.229	0	0.06	0.050	0.035	0.044	0	1	0.794	0.50	0.764	0
4	0.3	0.262	0.20	0.208	30	0.06	0.050	0.035	0.044	45	1	0.585	0.20	1.037	30
5	0.3	0.238	0.15	0.229	30	0.06	0.058	0.055	0.037	45	1	0.794	0.50	0.764	30
6	0.3	0.282	0.25	0.193	30	0.06	0.055	0.045	0.040	45	1	0.705	0.35	0.861	30
7	0.3	0.238	0.15	0.229	15	0.06	0.050	0.035	0.044	15	1	0.794	0.50	0.764	45
8	0.3	0.282	0.25	0.193	15	0.06	0.058	0.055	0.037	15	1	0.705	0.35	0.861	45
9	0.3	0.262	0.20	0.208	15	0.06	0.055	0.045	0.040	15	1	0.585	0.20	1.037	45
10	0.3	0.282	0.25	0.193	30	0.06	0.055	0.045	0.040	0	1	0.705	0.35	0.861	45
11	0.3	0.262	0.20	0.208	30	0.06	0.050	0.035	0.044	0	1	0.585	0.20	1.037	45
12	0.3	0.238	0.15	0.229	30	0.06	0.058	0.055	0.037	0	1	0.794	0.50	0.764	45
13	0.3	0.238	0.15	0.229	15	0.06	0.058	0.055	0.037	30	1	0.585	0.20	1.037	15
14	0.3	0.282	0.25	0.193	15	0.06	0.055	0.045	0.040	30	1	0.794	0.50	0.764	15
15	0.3	0.262	0.20	0.208	15	0.06	0.050	0.035	0.044	30	1	0.705	0.35	0.861	15
16	0.3	0.262	0.20	0.208	45	0.06	0.050	0.035	0.044	45	1	0.705	0.35	0.861	0
17	0.3	0.238	0.15	0.229	45	0.06	0.058	0.055	0.037	45	1	0.585	0.20	1.037	0
18	0.3	0.282	0.25	0.193	45	0.06	0.055	0.045	0.040	45	1	0.794	0.50	0.764	0
19	0.3	0.238	0.15	0.229	45	0.06	0.055	0.045	0.040	0	1	0.705	0.35	0.861	0
20	0.3	0.282	0.25	0.193	45	0.06	0.050	0.035	0.044	0	1	0.585	0.20	1.037	0
21	0.3	0.262	0.20	0.208	45	0.06	0.058	0.055	0.037	0	1	0.794	0.50	0.764	0
22	0.3	0.238	0.15	0.229	45	0.06	0.055	0.045	0.040	30	1	0.585	0.20	1.037	45
23	0.3	0.282	0.25	0.193	45	0.06	0.050	0.035	0.044	30	1	0.794	0.50	0.764	45
24	0.3	0.262	0.20	0.208	45	0.06	0.058	0.055	0.037	30	1	0.705	0.35	0.861	45
25	0.3	0.282	0.25	0.193	0	0.06	0.050	0.035	0.044	45	1	0.794	0.50	0.764	15
26	0.3	0.262	0.20	0.208	0	0.06	0.058	0.055	0.037	45	1	0.705	0.35	0.861	15
27	0.3	0.238	0.15	0.229	0	0.06	0.055	0.045	0.040	45	1	0.585	0.20	1.037	15
28	0.3	0.262	0.20	0.208	0	0.06	0.055	0.045	0.040	30	1	0.794	0.50	0.764	30
29	0.3	0.238	0.15	0.229	0	0.06	0.050	0.035	0.044	30	1	0.705	0.35	0.861	30
30	0.3	0.282	0.25	0.193	0	0.06	0.058	0.055	0.037	30	1	0.585	0.20	1.037	30
31	0.3	0.262	0.20	0.208	30	0.06	0.058	0.055	0.037	15	1	0.794	0.50	0.764	15
32	0.3	0.238	0.15	0.229	30	0.06	0.055	0.045	0.040	15	1	0.705	0.35	0.861	15
33	0.3	0.282	0.25	0.193	30	0.06	0.050	0.035	0.044	15	1	0.585	0.20	1.037	15
34	0.3	0.282	0.25	0.193	15	0.06	0.058	0.055	0.037	15	1	0.585	0.20	1.037	30
35	0.3	0.262	0.20	0.208	15	0.06	0.055	0.045	0.040	15	1	0.794	0.50	0.764	30
36	0.3	0.238	0.15	0.229	15	0.06	0.050	0.035	0.044	15	1	0.705	0.35	0.861	30

4.2.2 Model estimation

The first set of models estimated were RPL models with distributions that match the distributions used to generate the datasets (normal, triangular, uniform, lognormal and censored normal). For example, 3600 RPL models were estimated with normally distributed random parameters, on the datasets generated with normally distributed random parameters. Whilst in an empirical context the analyst does not know what the true distribution is (and indeed the true distribution is unlikely to assume a convenient distribution such as normal, triangular, etc.), the point of these simulations is to measure bias in the estimated distributions, which requires complete knowledge of the true distribution, including both its functional form and the moments describing it. Specifically, of interest is bias in the mean and dispersion of the random parameters. Across the five distributions, 18400 models were estimated. The censored normal models test the accuracy of the censored normal distribution in recovering ANA rates, and test for other biases that might be introduced. The RPL model was detailed in Section 2.1.3. Five thousand Halton draws were used for estimation. Section 3.3.2 provided a discussion as to why such a large number was employed.

The next set of model estimates seek to determine the accuracy of the IANA model, which was introduced in Section 2.2.2, and formalised in Section 3.2. Results from the CANA model are not presented, as preliminary testing revealed very similar results to the IANA model. The IANA model was estimated on four distributions, for a total of 29600 models. Since these models do not require integration across a large number of draws, estimation times are significantly less that the RPL models. As with the RPL model with censored normal distributions, the aim of estimating models on the simulated datasets with the latent class methods is to determine the accuracy with which ANA is estimated. However, the IANA model differs in that beyond ANA, no preference heterogeneity is estimated.

4.2.3 Modelling the impact of attribute nonattendance on choice model outputs

An inspection of the parameter values retrieved, and the implicit rates of sign violation, averaged across the 100 datasets per treatment, will allow for insights to be made into the impact of ANA on choice model results. Indeed, discussion based on such inspection will form a starting point for the analysis in this chapter. However, to more precisely quantify the impact, various models will be estimated, where the dependent variables are measures of bias, and proportions of sign violation.

Of interest is the bias in the moments describing the RP distributions. For the normal, triangular and uniform distributions, the structural parameters of the distribution also serve as the moments of the distribution, and so the biases in these parameters are considered. However, in the lognormal distribution, the structural parameters are of the underlying normal distribution, which is then exponentiated to obtain the coefficients. Bias in the structural parameters of the underlying normal distribution only has limited appeal. Of greater interest is bias in the moments and other measures of the lognormal distribution itself. This chapter will consider bias in four such measures: the mean ($e^{\mu+\sigma^2/2}$), the median (e^{μ}), the mode ($e^{\mu-\sigma^2}$), and the standard deviation ($\sqrt{(e^{\sigma^2}-1)e^{2\mu+\sigma^2}}$).

For each attribute k, we require a measurement of bias for each moment or measure h, where h may represent the mean, standard deviation, mode, etcetera, as detailed in the preceding paragraph. For any given treatment r, denote the true measure of the distribution used to generate the datasets as λ_{khr}. The estimated measure for dataset s is denoted as $\hat{\lambda}_{khrs}$, where this value will vary across the $S = 100$ datasets due to simulation noise. The measure of bias employed will be

$$Bias(\lambda_{khr}) = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{(\hat{\lambda}_{khrs} - \lambda_{khr})}{\lambda_{khr}}\right). \tag{4.3}$$

This is a relative measure of bias, which allows the observations from multiple parameters, with differing specifications, to be combined. For example, three means are employed with the normal distribution (-0.3, -0.06, and 0.9), and the bias in the mean can be modelled with observations from all three parameters. Also, the formula effectively cancels out the sign of the measure, allowing a focus on the change in *magnitude* of the measure. Consider again the bias in the mean. It is hypothesised that nonattendance to an attribute will bias the magnitude of the mean sensitivity to that parameter downwards, closer to zero. With the above formula, both a change in parameter one from -0.3 to -0.15 and a change in parameter three from 0.9 to 0.45 have the same bias of -0.5, which represents a reduction in magnitude of 50 percent. The bias measurement could be multiplied by 100 to obtain a percentage, but will remain as a proportion to be consistent with some of the explanatory variables introduced into the models, which will be detailed shortly.

To model the bias in the measures, a series of regression models were estimated. The bias measure introduced in Formula 4.3 serves as the dependent variable. Independent variables include the true rate of ANA, and various transformations of and interactions with the true coefficient of variation (C.V). The coefficient of variation, defined as the standard deviation divided by the mean, provides a normalised measure of dispersion, which is useful in that it allows observations from all three parameters to be stacked and used in the one model (where this also motivates the use of a relative measure of bias, as discussed above). The various transformations and interactions will be introduced as each of the specific models estimated are introduced, together with specific justifications. For the normal, triangular, uniform and censored normal distributions, two regressions were estimated - the bias in the mean and in the standard deviation/spread. For the lognormal distribution, five regressions were estimated, for the mean, median, mode, standard deviation and 99$^{\text{th}}$ percentile of the lognormal distribution itself (not the underlying normal distribution). Separate estimation of these regressions ignores the possibility that the errors could be correlated. Consequently, it was appropriate to estimate models such as seemingly unrelated regression equation (SURE) models, with five regression equations for the lognormal distribution, and two regression equations for the other distributions.

Modelling the proportions of sign violation for the normal, triangular and uniform distributions, and the point mass at zero for the censored normal distribution, requires a model that can handle the dependent variable being bounded between the limits of zero and one. The truncated regression model is one candidate, however, it omits observations that do not lie between the limits, which is not appropriate in this context, where the proportion of sign violations might legitimately be zero. More appropriate, and applied in this chapter, is the two-limit Tobit model, which estimates a latent regression, but censors the dependent variable to be at the limit, should it be exceeded in the latent regression. Evaluating model fit is complicated by the nonlinearity of the model, but the $\rho^2_{DECOMPOSITION}$ measure provided in Limdep and detailed in Greene (2002) has been used, and is referred to merely as ρ^2 in the tables and discussion, for the sake of brevity.

The IANA model only provides a point estimate for the sensitivity to each attribute, and does not capture preference heterogeneity. Since only a single regression equation is thus estimated, a simple regression model is estimated in place of the SURE model. Since the measure of respondents that exhibit ANA is a proportion, a two-limit Tobit model is employed for this dependent variable. Table 4.4 summarises the models estimated for each dependent variable for each combination of choice model and underlying distribution examined.

4.3 Random parameters logit model results

Sections 4.3 and 4.4 present the detailed results from the simulations. Section 4.5 will summarise the findings, and discuss them further.

Table 4.4: Models employed for each dependent variable

Model	Data generation	Mean	Std dev.	Spread	Median	Mode	99^{th} per.	Sign violation	ANA/ Zero mass
RPL	Normal	SURE	SURE	-	-	-	-	Tobit	-
RPL	Triangular	SURE	-	SURE	-	-	-	Tobit	-
RPL	Uniform	SURE	-	SURE	-	-	-	Tobit	-
RPL	Lognormal	SURE	SURE	-	SURE	SURE	SURE	-	-
RPL	Censored normal	SURE	SURE	-	-	-	-	-	Tobit
IANA	Normal	Regress.	-	-	-	-	-	-	Tobit
IANA	Triangular	Regress.	-	-	-	-	-	-	Tobit
IANA	Uniform	Regress.	-	-	-	-	-	-	Tobit
IANA	Lognormal	Regress.	-	-	-	-	-	-	Tobit

4.3.1 Normal distribution

Table 4.5 presents the results for the models estimated on the datasets generated with normally distributed random parameters. The true parameter means listed in the footnotes are fixed across treatments for each of the three parameters, and so only the true standard deviations and rates of ANA (labelled NA for brevity) are presented for each treatment to allow ready comparison with the estimated values. Estimated means, standard deviations and percentages of sign violation as presented in the table are averages over the 100 datasets per treatment. Where parameters have a zero rate of ANA, the true mean and standard deviation values are retrieved adequately, with no obvious signs of bias.

Where ANA rates are greater than zero, a clear downward bias in the magnitude of the parameter means is evident across all three parameters, with the extent of the bias increasing with the rate of ANA. The true standard deviation appears to have some impact on the magnitude of the estimated mean, primarily for parameters two and three, with greater downward bias for the high dispersion parameter specifications.

The impact of ANA on the estimated standard deviation is heavily influenced by the true dispersion of the random parameter. For the low dispersion specification ($\sigma = 0.1, 0.02, 0.3$ for parameters one, two and three respectively), there is a large upwards bias in the standard deviation. For the moderate dispersion specification, there is a slight upwards bias, and for the high dispersion specification, there is a small downwards bias. The influence of the rate of ANA does not appear to be strong, with the range of estimated standard deviations appearing to be slightly compressed across the three dispersion specifications as the ANA rate increases. However, by far the greater impact is the bias in the estimated standard deviations if there is some degree of nonattendance, relative to an absence of bias with full attendance. This suggests that the true state of preference heterogeneity may have a dominant influence on the extent of bias induced by ANA.

Assessing the impact of ANA on the estimated percentage of sign violation when the true distribution is normally distributed is complicated by the true distribution itself inducing sign violation. When inspecting the percentages in Table 4.5, it is useful to consider the true percentage. Since the same coefficient of variation is applied across all three parameters, the true percentages are the same for the low, medium and high dispersion specifications regardless of parameter, with percentages of 0.1, 6.7 and 15.9 percent respectively. A near zero true percentage for the low dispersion specification means that higher percentages are almost entirely induced by ANA. An inspection of the low dispersion specifications shows that indeed ANA does lead to sign violation, with more violations as the ANA rate increases. Attribute nonattendance also increases the percentage of sign violation beyond the true value for all dispersion specifications. Further, holding the ANA rate constant, the difference in percentage of sign violations between the low and high dispersion conditions appears to be roughly equal to the difference of the true percentages. This suggests that ANA is probably not interacting with preference heterogeneity in its influence over the percentage of sign violations. In summary, there is a clear link between ANA and sign violations, with increases in ANA increasing the incidence of sign violation.

Table 4.6 presents the model results. The model fits are strong, with the ρ^2 of the bias in the mean particularly high at 0.96999. Unlike in Table 4.5, the ANA rate entered into the models is expressed as a proportion. The 'ANA rate' parameter is highly significant, with a value close to minus one. This suggests that the downward bias in the mean of a normally distributed parameter is a percentage that is close to, but slightly higher than the true percentage of the sample that does not attend to the corresponding attribute. For example, an ANA rate of 0.3 (i.e., 30 percent of the sample) translates to a bias measure of -0.3264, ceteris paribus, which is a decrease in the magnitude of the mean of 32.64 percent. The inverse of the coefficient of variation, multiplied by a dummy variable set to one if the true ANA rate is greater than zero ('(1/C.V.)$\times$dummy$_{\text{ANA}>0}$'), is also entered into the regression equation of the bias in the mean. Irrespective of which of the three parameters, in the presence of nonzero ANA this variable takes the value of three for the low dispersion specification (e.g., $\frac{1}{(0.2/0.6)}$ for parameter one), and 1.5 and one for the medium and high dispersion specifications respectively. The positively signed parameter for this variable in the mean regression suggests that the smaller the dispersion, the larger the estimate of the dependent variable. In the

Table 4.5: RPL models estimated on datasets with normally distributed true sensitivities

Tr.	**Parameter One**					**Parameter Two**					**Parameter Three**				
	Actual[1]		**Estimated**			**Actual**[2]		**Estimated**			**Actual**[3]		**Estimated**		
	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**>0** (%)	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**>0** (%)	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**<0** (%)
1	0.1	0	-0.300	0.097	0.1	0.02	0	-0.061	0.023	0.5	0.6	0	0.932	0.605	6.2
2	0.2	0	-0.314	0.199	5.7	0.04	0	-0.058	0.040	7.3	0.9	0	0.894	0.891	15.8
3	0.3	0	-0.321	0.297	14.0	0.06	0	-0.058	0.059	16.5	0.3	0	0.901	0.308	0.3
4	0.2	30	-0.218	0.216	15.6	0.06	45	-0.029	0.052	29.0	0.9	30	0.606	0.816	22.9
5	0.3	30	-0.220	0.271	20.8	0.02	45	-0.032	0.033	16.9	0.3	30	0.633	0.492	9.9
6	0.1	30	-0.218	0.171	10.2	0.04	45	-0.031	0.041	22.8	0.6	30	0.616	0.626	16.2
7	0.3	15	-0.266	0.280	17.1	0.06	15	-0.046	0.058	21.3	0.3	45	0.484	0.489	16.1
8	0.1	15	-0.262	0.150	4.0	0.02	15	-0.050	0.028	3.8	0.6	45	0.466	0.595	21.6
9	0.2	15	-0.264	0.211	10.5	0.04	15	-0.048	0.042	12.7	0.9	45	0.449	0.740	27.2
10	0.1	30	-0.217	0.171	10.2	0.04	0	-0.058	0.039	7.1	0.6	45	0.467	0.598	21.7
11	0.2	30	-0.217	0.215	15.6	0.06	0	-0.057	0.060	17.2	0.9	45	0.449	0.741	27.2
12	0.3	30	-0.221	0.281	21.5	0.02	0	-0.059	0.024	0.8	0.3	45	0.491	0.497	16.1
13	0.3	15	-0.267	0.284	17.3	0.02	30	-0.041	0.033	10.4	0.9	15	0.758	0.864	19.0
14	0.1	15	-0.264	0.147	3.7	0.04	30	-0.040	0.043	17.8	0.3	15	0.766	0.423	3.6
15	0.2	15	-0.265	0.209	10.2	0.06	30	-0.038	0.056	24.9	0.6	15	0.763	0.627	11.2
16	0.2	45	-0.168	0.207	20.9	0.06	45	-0.029	0.052	29.0	0.6	0	0.902	0.603	6.8
17	0.3	45	-0.165	0.248	25.3	0.02	45	-0.032	0.033	16.5	0.9	0	0.901	0.901	15.8
18	0.1	45	-0.169	0.175	16.7	0.04	45	-0.031	0.041	23.0	0.3	0	0.904	0.307	0.3
19	0.3	45	-0.164	0.250	25.5	0.04	0	-0.058	0.040	7.5	0.6	0	0.902	0.605	6.8
20	0.1	45	-0.168	0.173	16.5	0.06	0	-0.057	0.060	16.9	0.9	0	0.894	0.894	15.9
21	0.2	45	-0.166	0.212	21.6	0.02	0	-0.059	0.023	0.7	0.3	0	0.903	0.307	0.3
22	0.3	45	-0.163	0.244	25.3	0.04	30	-0.039	0.043	18.2	0.9	45	0.450	0.740	27.1
23	0.1	45	-0.169	0.172	16.4	0.06	30	-0.038	0.056	25.0	0.3	45	0.485	0.492	16.2
24	0.2	45	-0.166	0.204	20.8	0.02	30	-0.041	0.033	10.6	0.6	45	0.471	0.599	21.6
25	0.1	0	-0.308	0.100	0.1	0.06	45	-0.029	0.053	29.1	0.3	15	0.764	0.423	3.6
26	0.2	0	-0.314	0.199	5.8	0.02	45	-0.032	0.033	16.7	0.6	15	0.763	0.624	11.1
27	0.3	0	-0.319	0.299	14.3	0.04	45	-0.031	0.041	23.0	0.9	15	0.754	0.856	18.9
28	0.2	0	-0.314	0.198	5.7	0.04	30	-0.040	0.043	17.8	0.3	30	0.626	0.483	9.7
29	0.3	0	-0.319	0.298	14.2	0.06	30	-0.038	0.056	24.8	0.6	30	0.612	0.623	16.3
30	0.1	0	-0.307	0.103	0.2	0.02	30	-0.041	0.032	10.0	0.9	30	0.608	0.828	23.1
31	0.2	30	-0.219	0.216	15.5	0.02	15	-0.050	0.028	4.1	0.3	15	0.772	0.435	3.9
32	0.3	30	-0.219	0.272	21.0	0.04	15	-0.048	0.042	12.7	0.6	15	0.766	0.636	11.4
33	0.1	30	-0.217	0.171	10.2	0.06	15	-0.047	0.058	21.0	0.9	15	0.759	0.864	19.0
34	0.1	15	-0.262	0.150	4.0	0.02	15	-0.050	0.027	3.7	0.9	30	0.607	0.821	23.0
35	0.2	15	-0.265	0.208	10.2	0.04	15	-0.048	0.042	12.6	0.3	30	0.626	0.485	9.9
36	0.3	15	-0.266	0.282	17.3	0.06	15	-0.046	0.058	21.2	0.6	30	0.614	0.628	16.4

1. μ =-0.3 for all treatments. 2. μ =-0.06 for all treatments. 3. μ =0.9 for all treatments.

context of estimates that are negative, as in this dataset, this implies that a smaller dispersion will lead to less downward bias in the mean. This is consistent with the inspection of the data presented earlier.

Table 4.6: SURE and Tobit model results - RPL models estimated on datasets with normally distributed true sensitivities

	μ		σ		**Sign violation**	
	Par.	***t*-ratio**	**Par.**	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	-1.088	-55.17			0.321	40.42
(1/C.V.)×dummy$_{\text{ANA}>0}$	0.008	2.63	0.345	36.69		
Dummy$_{\text{ANA}>0}$			-0.452	-23.81		
True sign violation					0.855	41.42
Constant					0.007	2.4
ρ^2	0.96999		0.93067		0.43442	

The '(1/C.V.)×dummy$_{\text{ANA}>0}$' variable has a significant impact on the bias of the standard deviation. The positive parameter needs to be interpreted in conjunction with the negative 'dummy$_{\text{ANA}>0}$' parameter. The high dispersion specification in the dataset will lead to 1/C.V. evaluating to one, and the estimate evaluating to $0.345 - 0.452 = -0.107$, which is a decrease of 10.7 percent. The medium and low dispersion specifications evaluate to increases of 6.6 percent and 58.3 percent respectively, again a pattern consistent with the inspection of the data.

The Tobit model for sign violation is also presented in Table 4.6. A comparison of ρ^2 to the SURE regressions is meaningless[6], but the ρ^2 is useful in comparing alternative Tobit model specifications. The coefficients estimated are associated with the underlying latent regression, and should not be interpreted directly. Instead, the marginal effects can be computed, which for the Tobit model are the regression coefficients multiplied by the probability that the observation is not censored. However, due presumedly to a very low probability of being censored, the marginal effects have the same values as the regression coefficients, and so the values reported can be interpreted as either regression coefficients or marginal effects. Both the true sign violations and ANA rates (entered into the model as proportions) have a highly significant effect. Approximately one third of the ANA rate is translated into estimated sign violation, and with a marginal effect of 0.855, true sign violation is largely but not completely captured by the estimated sign violation. Notably, the true dispersion of the parameter does not interact with ANA, and only has an impact on estimated sign violation through the

[6] A ρ^2 of 0.98934 for a regression model estimated with the same parameters using ordinary least squares provides some clue that the model fit is very good, but this result should be treated with caution.

true sign violation, which is itself a function of the dispersion of the normally distributed random parameter. The probability that each observation has not been censored to the limit values can be computed. Ninety nine of the 108 probabilities are 99.99 percent or greater. The remaining nine are all associated with zero percent ANA, and low dispersion. For these observations, there is no biasing influence of ANA, and a true sign violation of 0.1 percent, and so the higher probability of being censored is plausible.

The key findings for the normal distribution are that the mean is biased down by a percentage roughly equal to the ANA rate; the bias in the standard deviation may be upwards or downwards depending on the true measure of dispersion; and the percentage of coefficients that violate sign is increased by about a third of the nonattendance percentage.

4.3.2 Triangular distribution

Table 4.7 presents the results for the choice models estimated on the datasets generated using the triangularly distributed random parameters. Under parameter specifications with full attendance, the true parameter estimates are retrieved accurately, with the exception being σ under the low dispersion specifications for parameters two and three, where in both cases the estimated values are overestimates of the true values. The bias in the magnitude of the mean due to ANA is consistently downwards, and is proportional to the true ANA rate. There is an upward bias in the spread parameter, σ, although the bias appears to be disproportionately large in percentage terms for the low dispersion parameter specifications, which hints at a possible non-linearity in the bias. The extent of sign violation, which is not present in the triangular distributions used to generate sensitivities for those who attend to the attribute, increases as the ANA rate increases. There is also an increase in sign violation associated with an increasing levels of dispersion in the true distribution.

Table 4.8 contains the SURE and Tobit model results for the triangular distribution. As with the normal distribution, the percentage of respondents that do not attend to an attribute leads to roughly the same percentage decrease in the magnitude of the mean (due to a parameter value of -0.998). The bias in the measure of dispersion (in this case the spread) is again positively influenced by the inverse of the coefficient of variation, although in this case this value is first squared ($(1/\text{C.V.})^2$) before being multiplied by '$\text{dummy}_{\text{ANA}>0}$'. This captures the observed non-linearity in the bias. Also, whereas with the normal distribution the constant was only entered for parameter specifications that contained some degree of ANA, under the triangular specification the constant is present for all parameter specifications. This

Table 4.7: RPL models estimated on datasets with triangularly distributed true sensitivities

Tr.	Parameter One					Parameter Two					Parameter Three				
	Actual[1]		Estimated			Actual[2]		Estimated			Actual[3]		Estimated		
	σ	NA (%)	$\bar{\mu}$	$\bar{\sigma}$	>0 (%)	σ	NA (%)	$\bar{\mu}$	$\bar{\sigma}$	>0 (%)	σ	NA (%)	$\bar{\mu}$	$\bar{\sigma}$	<0 (%)
1	0.1	0	-0.302	0.109	0.0	0.02	0	-0.061	0.037	0.0	0.6	0	0.906	0.603	0.0
2	0.2	0	-0.306	0.198	0.0	0.04	0	-0.059	0.038	0.0	0.9	0	0.896	0.877	0.2
3	0.3	0	-0.308	0.295	0.0	0.06	0	-0.060	0.061	0.5	0.3	0	0.902	0.476	0.0
4	0.2	30	-0.217	0.398	10.3	0.06	45	-0.032	0.084	19.1	0.9	30	0.625	1.235	12.2
5	0.3	30	-0.220	0.431	12.0	0.02	45	-0.033	0.075	15.3	0.3	30	0.631	1.061	8.2
6	0.1	30	-0.215	0.375	9.0	0.04	45	-0.033	0.078	16.8	0.6	30	0.626	1.120	9.7
7	0.3	15	-0.263	0.389	5.3	0.06	15	-0.050	0.075	5.7	0.3	45	0.490	1.113	15.6
8	0.1	15	-0.259	0.307	1.3	0.02	15	-0.051	0.053	1.0	0.6	45	0.485	1.155	16.8
9	0.2	15	-0.260	0.342	2.9	0.04	15	-0.050	0.063	2.7	0.9	45	0.481	1.233	18.5
10	0.1	30	-0.215	0.375	9.1	0.04	0	-0.059	0.035	0.0	0.6	45	0.487	1.166	16.9
11	0.2	30	-0.217	0.398	10.3	0.06	0	-0.059	0.059	0.6	0.9	45	0.483	1.242	18.6
12	0.3	30	-0.216	0.427	12.3	0.02	0	-0.060	0.040	0.0	0.3	45	0.496	1.140	15.9
13	0.3	15	-0.263	0.389	5.2	0.02	30	-0.042	0.070	8.0	0.9	15	0.764	1.124	5.2
14	0.1	15	-0.260	0.300	1.0	0.04	30	-0.042	0.074	9.6	0.3	15	0.764	0.819	0.8
15	0.2	15	-0.261	0.336	2.5	0.06	30	-0.041	0.083	12.8	0.6	15	0.762	0.947	2.2
16	0.2	45	-0.170	0.410	17.1	0.06	45	-0.032	0.084	19.0	0.6	0	0.901	0.579	0.0
17	0.3	45	-0.171	0.435	18.4	0.02	45	-0.033	0.074	15.3	0.9	0	0.902	0.894	0.2
18	0.1	45	-0.168	0.392	16.3	0.04	45	-0.033	0.079	16.8	0.3	0	0.903	0.460	0.0
19	0.3	45	-0.171	0.437	18.6	0.04	0	-0.059	0.037	0.0	0.6	0	0.902	0.581	0.0
20	0.1	45	-0.169	0.389	16.1	0.06	0	-0.059	0.058	0.5	0.9	0	0.900	0.876	0.2
21	0.2	45	-0.168	0.409	17.3	0.02	0	-0.060	0.037	0.0	0.3	0	0.906	0.424	0.0
22	0.3	45	-0.171	0.429	18.1	0.04	30	-0.041	0.076	10.3	0.9	45	0.486	1.248	18.6
23	0.1	45	-0.169	0.388	15.9	0.06	30	-0.041	0.083	12.9	0.3	45	0.491	1.128	15.9
24	0.2	45	-0.170	0.405	16.8	0.02	30	-0.042	0.071	8.2	0.6	45	0.488	1.169	16.9
25	0.1	0	-0.303	0.115	0.0	0.06	45	-0.032	0.085	19.1	0.3	15	0.759	0.820	0.7
26	0.2	0	-0.306	0.192	0.0	0.02	45	-0.033	0.075	15.5	0.6	15	0.763	0.955	2.2
27	0.3	0	-0.309	0.297	0.0	0.04	45	-0.032	0.078	16.9	0.9	15	0.764	1.130	5.3
28	0.2	0	-0.305	0.191	0.0	0.04	30	-0.041	0.075	10.1	0.3	30	0.626	1.041	7.9
29	0.3	0	-0.308	0.295	0.0	0.06	30	-0.041	0.083	12.9	0.6	30	0.626	1.116	9.6
30	0.1	0	-0.303	0.120	0.0	0.02	30	-0.042	0.069	7.7	0.9	30	0.627	1.232	12.0
31	0.2	30	-0.217	0.394	10.2	0.02	15	-0.051	0.054	1.0	0.3	15	0.765	0.846	1.0
32	0.3	30	-0.219	0.431	12.1	0.04	15	-0.050	0.063	2.6	0.6	15	0.766	0.976	2.6
33	0.1	30	-0.215	0.372	8.9	0.06	15	-0.050	0.075	5.6	0.9	15	0.764	1.128	5.2
34	0.1	15	-0.259	0.304	1.2	0.02	15	-0.051	0.054	1.2	0.9	30	0.623	1.226	12.1
35	0.2	15	-0.261	0.337	2.6	0.04	15	-0.050	0.062	2.4	0.3	30	0.627	1.045	8.0
36	0.3	15	-0.263	0.388	5.2	0.06	15	-0.050	0.075	5.7	0.6	30	0.625	1.120	9.7

1. μ =-0.3 for all treatments. 2. μ =-0.06 for all treatments. 3. μ =0.9 for all treatments.

is likely due to the aforementioned upwards bias in σ for some treatments under complete attribute attendance.

The Tobit model of the proportion of sign violation again has marginal effects that equal the coefficients, and so only the former are reported in Table 4.8. The ANA rate has a very significant positive influence on sign violation. The '$(1/\text{C.V.})\times\text{dummy}_{\text{ANA}>0}$' parameter is negative and highly significant, which means that as the spread (i.e., the preference heterogeneity) decreases, so too does the rate of sign violation, consistent with the earlier inspection of the data. The probabilities that each observation has not been censored again evaluate to 99.99 percent or above, with the exceptions of observations with full attendance (27 percent), and observations with 15 percent ANA and low dispersion (93.79 percent). A lack of sign violation under full attendance is to be expected for the triangular distribution, given that there is no such violation in the true distribution.

Table 4.8: SURE and Tobit model results - RPL models estimated on datasets with triangularly distributed true sensitivities

	μ		σ		**Sign violation**	
	Par.	***t*-ratio**	**Par.**	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	-0.998	-210.11			0.462	87.5
$(1/\text{C.V.})^2\times\text{dummy}_{\text{ANA}>0}$			0.249	31.71		
$(1/\text{C.V.})\times\text{dummy}_{\text{ANA}>0}$					-0.018	-21.96
Constant			0.167	4.55	-0.0045	-3.13
ρ^2	0.99332		0.90178		0.34954	

The findings with the triangular distribution are broadly similar to the normal distribution. Downward bias in the mean is again roughly the same as the ANA rate, in percentage terms. However, the true dispersion has no impact on bias in the mean for the triangular distribution, unlike for the normal distribution. This might be due to the lack of sign violations in the underlying datasets generated for the triangular distribution. The magnitude of the true dispersion in the normal distribution will influence the extent of true sign violation, which in turn will lead to some coefficients of different sign to the mean. This in turn may impact the estimated mean. While the specification of the sign violation model is different to the normal distribution, the results are basically the same. The sign violation will increase as either the ANA rate or true dispersion increase. For the normal distribution, the impact of the dispersion is captured by the true sign violation variable (which increases with the dispersion), whereas for the triangular distribution it is captured by a negative parameter estimate for '$(1/\text{C.V.})\times\text{dummy}_{\text{ANA}>0}$'.

4.3.3 Uniform distribution

The choice model outputs for the datasets generated with uniformly distributed random parameters are detailed in Table 4.9. An inspection of the values suggests that the biases induced by ANA on the uniform distribution are fairly similar to those on the triangular distribution. Table 4.10 contains the results from the SURE and Tobit models for the uniform distribution.

As with the triangular distribution, ANA biases the mean of the random parameter downwards, by the same percentage as the ANA rate, with no other influence evident. This is reflected in the corresponding regression equation, which contains the ANA rate as a single regressor, with a highly significant parameter value of -0.994 and a ρ^2 of 0.98237.

The bias in the spread parameter is only slightly influenced by ANA, and the percentage bias is considerably greater for specifications with low true dispersion. As with the triangular distribution, this pattern is captured in the regression by a positive parameter associated with '$(1/\text{C.V.})^2 \times \text{dummy}_{\text{ANA}>0}$'. Unlike the triangular distribution, there is no constant in the regression equation, likely due to no obvious upwards bias in the spread under parameter specifications with full attendance.

The rate of sign violation is influenced by the ANA rate, and, so long as the ANA rate is greater than zero, sign violations increase as the true dispersion increases. Table 4.10 contains the Tobit model, which has the same specification as the triangular distribution and consistent results. Also consistent with the triangular distribution are the probabilities that each observation has not been censored. These evaluate to 99.99 percent or above for all observations, with the exceptions of observations with full attendance (50 percent), and observations with 15 percent ANA and low dispersion (57.35 percent). A comparison between the triangular and uniform distributions of sign violation rates, from Tables 4.7 and 4.9 respectively, reveals higher rates for the uniform distribution. While the means and spreads are the same for each treatment, easing comparison, such a comparison must still be made with caution. Consider the case where the mean equals the spread. The triangular distribution will taper down from a peak at the mean to zero. The uniform distribution will remain flat from the mean to zero. Consequently, the uniform distribution will have more coefficients that are close to zero. Any bias downwards in the mean or upwards in the spread will likely move disproportionately more mass to the other sign.

Table 4.9: RPL models estimated on datasets with uniformly distributed true sensitivities

Tr.	**Parameter One**					**Parameter Two**					**Parameter Three**				
	Actual[1]		**Estimated**			**Actual**[2]		**Estimated**			**Actual**[3]		**Estimated**		
	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**>0** (%)	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**>0** (%)	σ	**NA** (%)	$\bar{\mu}$	$\bar{\sigma}$	**<0** (%)
1	0.1	0	-0.302	0.095	0.0	0.02	0	-0.060	0.024	0.0	0.6	0	0.911	0.596	0.0
2	0.2	0	-0.307	0.200	0.0	0.04	0	-0.059	0.039	0.0	0.9	0	0.888	0.880	1.3
3	0.3	0	-0.311	0.298	0.1	0.06	0	-0.059	0.059	1.8	0.3	0	0.898	0.320	0.0
4	0.2	30	-0.218	0.301	13.7	0.06	45	-0.032	0.065	25.7	0.9	30	0.626	0.986	18.1
5	0.3	30	-0.227	0.343	16.8	0.02	45	-0.033	0.053	19.0	0.3	30	0.627	0.770	9.0
6	0.1	30	-0.212	0.269	10.4	0.04	45	-0.032	0.058	22.0	0.6	30	0.621	0.850	13.2
7	0.3	15	-0.265	0.330	9.7	0.06	15	-0.049	0.065	12.1	0.3	45	0.487	0.804	19.5
8	0.1	15	-0.256	0.221	0.0	0.02	15	-0.050	0.041	0.4	0.6	45	0.482	0.856	21.7
9	0.2	15	-0.259	0.270	2.3	0.04	15	-0.050	0.051	3.2	0.9	45	0.481	0.949	24.6
10	0.1	30	-0.212	0.268	10.5	0.04	0	-0.059	0.039	0.0	0.6	45	0.485	0.862	21.8
11	0.2	30	-0.218	0.301	13.8	0.06	0	-0.059	0.060	2.6	0.9	45	0.484	0.956	24.6
12	0.3	30	-0.229	0.348	17.0	0.02	0	-0.060	0.026	0.0	0.3	45	0.495	0.806	19.1
13	0.3	15	-0.266	0.333	10.0	0.02	30	-0.042	0.051	8.6	0.9	15	0.762	0.972	10.6
14	0.1	15	-0.256	0.214	0.0	0.04	30	-0.041	0.057	13.5	0.3	15	0.758	0.603	0.0
15	0.2	15	-0.260	0.266	1.7	0.06	30	-0.041	0.068	19.9	0.6	15	0.758	0.768	2.1
16	0.2	45	-0.175	0.300	20.7	0.06	45	-0.032	0.065	25.6	0.6	0	0.898	0.602	0.0
17	0.3	45	-0.184	0.334	22.5	0.02	45	-0.033	0.054	19.2	0.9	0	0.898	0.900	1.5
18	0.1	45	-0.169	0.280	19.8	0.04	45	-0.032	0.060	22.8	0.3	0	0.902	0.314	0.0
19	0.3	45	-0.185	0.336	22.5	0.04	0	-0.059	0.039	0.0	0.6	0	0.901	0.602	0.0
20	0.1	45	-0.169	0.277	19.6	0.06	0	-0.059	0.059	2.3	0.9	0	0.892	0.885	1.3
21	0.2	45	-0.177	0.305	21.0	0.02	0	-0.060	0.025	0.0	0.3	0	0.906	0.318	0.0
22	0.3	45	-0.183	0.330	22.2	0.04	30	-0.041	0.059	14.7	0.9	45	0.490	0.959	24.4
23	0.1	45	-0.169	0.277	19.3	0.06	30	-0.041	0.068	19.7	0.3	45	0.488	0.803	19.4
24	0.2	45	-0.175	0.298	20.6	0.02	30	-0.042	0.052	9.2	0.6	45	0.488	0.865	21.7
25	0.1	0	-0.304	0.093	0.0	0.06	45	-0.032	0.066	25.7	0.3	15	0.759	0.605	0.0
26	0.2	0	-0.308	0.198	0.0	0.02	45	-0.033	0.053	18.9	0.6	15	0.760	0.770	2.1
27	0.3	0	-0.312	0.303	0.4	0.04	45	-0.032	0.058	22.1	0.9	15	0.760	0.971	10.7
28	0.2	0	-0.307	0.197	0.0	0.04	30	-0.041	0.057	13.8	0.3	30	0.622	0.751	8.4
29	0.3	0	-0.311	0.301	0.3	0.06	30	-0.040	0.068	20.0	0.6	30	0.620	0.858	13.6
30	0.1	0	-0.304	0.104	0.0	0.02	30	-0.042	0.050	8.3	0.9	30	0.630	1.006	18.6
31	0.2	30	-0.217	0.297	13.5	0.02	15	-0.050	0.040	0.2	0.3	15	0.762	0.625	0.1
32	0.3	30	-0.227	0.344	17.0	0.04	15	-0.050	0.051	3.3	0.6	15	0.763	0.785	2.6
33	0.1	30	-0.212	0.267	10.3	0.06	15	-0.049	0.065	11.8	0.9	15	0.759	0.962	10.3
34	0.1	15	-0.256	0.221	0.0	0.02	15	-0.051	0.040	0.4	0.9	30	0.623	0.990	18.4
35	0.2	15	-0.259	0.265	1.7	0.04	15	-0.050	0.051	2.9	0.3	30	0.621	0.754	8.6
36	0.3	15	-0.265	0.330	9.8	0.06	15	-0.049	0.065	12.2	0.6	30	0.621	0.858	13.6

1. μ =-0.3 for all treatments. 2. μ =-0.06 for all treatments. 3. μ =0.9 for all treatments.

Table 4.10: SURE and Tobit model results - RPL models estimated on datasets with uniformly distributed true sensitivities

	μ		σ		**Sign violation**	
	Par.	***t*-ratio**	**Par.**	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	-0.994	-130.07			0.577	31.61
$(1/\text{C.V.})^2 \times \text{dummy}_{\text{ANA}>0}$			0.163	49.72		
$(1/\text{C.V.}) \times \text{dummy}_{\text{ANA}>0}$					-0.037	-13.92
$\text{Dummy}_{\text{ANA}>0}$					0.027	3.57
ρ^2	0.98237		0.92787		0.34930	

4.3.4 Lognormal distribution

The impact of ANA on the accuracy of the recovery of the lognormal distribution is considered herein with respect to the mean and standard deviation of the lognormal distribution, rather than the underlying normal distribution. Since the lognormal distribution is asymmetrical, the mode and median are not necessarily the same as the mean, and bias in these two measures will also be examined. Finally, the lognormal distribution has a long tail, and any bias in the tail could have a pronounced impact on the behavioural interpretation of the model (for example, a longer tail could result in some WTP measures being more extreme). Bias in the 99^{th} percentile will be investigated to provide insight into the impact of ANA on the tail of the distribution. The lognormal distribution is strictly positive, and so there is no sign violation to consider[7]. As detailed previously, the parameter specifications fixed the mean and varied the standard deviation of the lognormal distribution. This also leads to variation in the true median and mode, with both values decreasing as the standard deviation increases.

The inspection of the lognormal results will begin with plots of the estimated lognormal distributions for each of the three parameters, in Figures 4.2, 4.3 and 4.4. Each figure contains three plots, for the low, medium and high dispersion dispersion specifications. Within each plot, there are four distributions plotted, for each of the four ANA rates specified. The full attendance distribution is specified with the true values of μ and σ, while the three nonattendance distributions are specified with the average values of the estimated μ and σ. The average is across the values retrieved from the three treatments that have the corresponding dispersion specification and ANA rate, detailed in Tables 4.11, 4.12 and 4.13.

The impact of ANA is clear from the figures: for any given parameter and dispersion specification, an increase in the ANA rate leads to a decrease in the mode of the lognormal

[7]Negative parameters are modelled by reversing the sign of the attributes for estimation, and then reversing the sign of the measures of central tendency for interpretation.

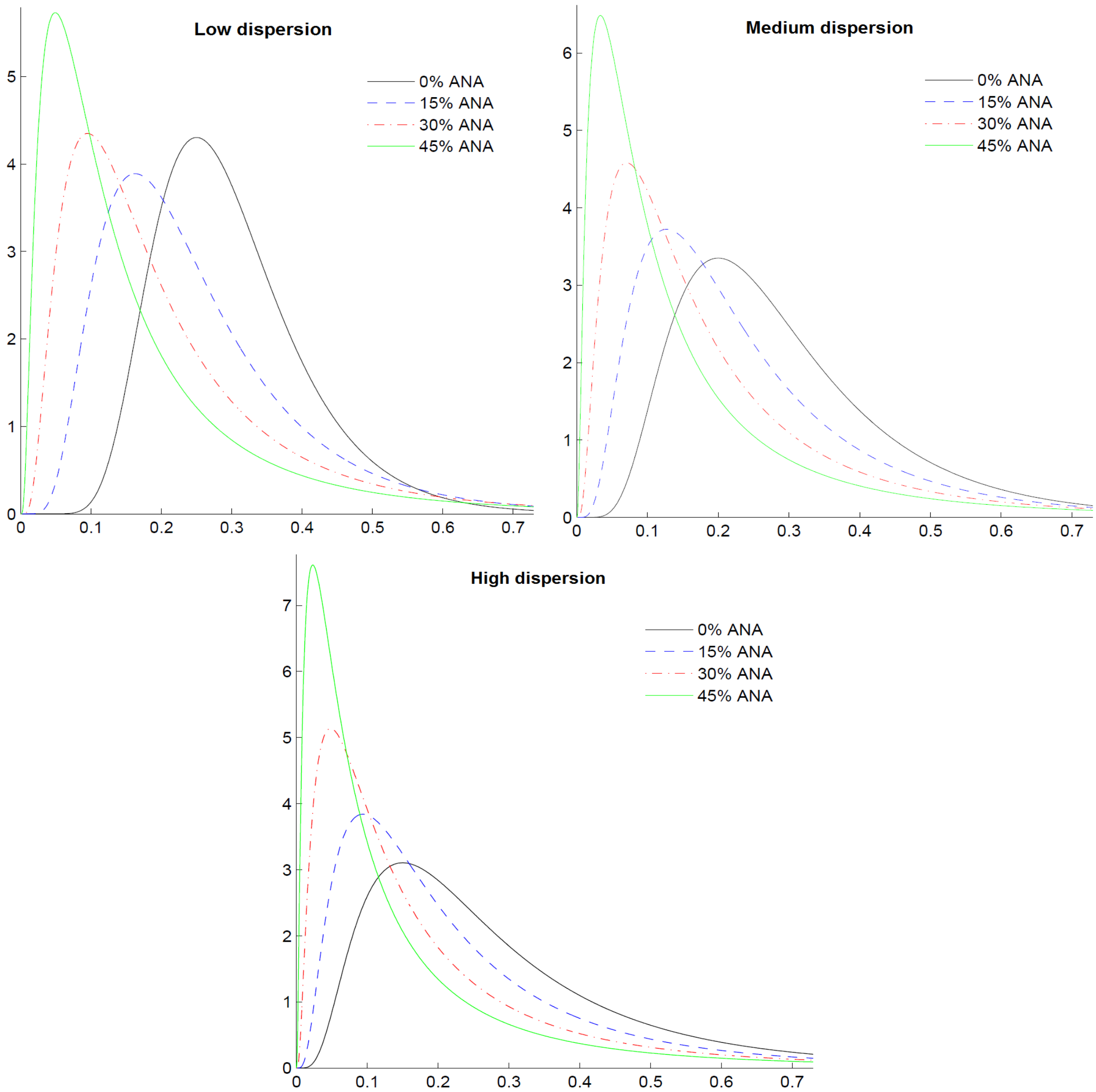

Figure 4.2: Estimated lognormal distributions for parameter one

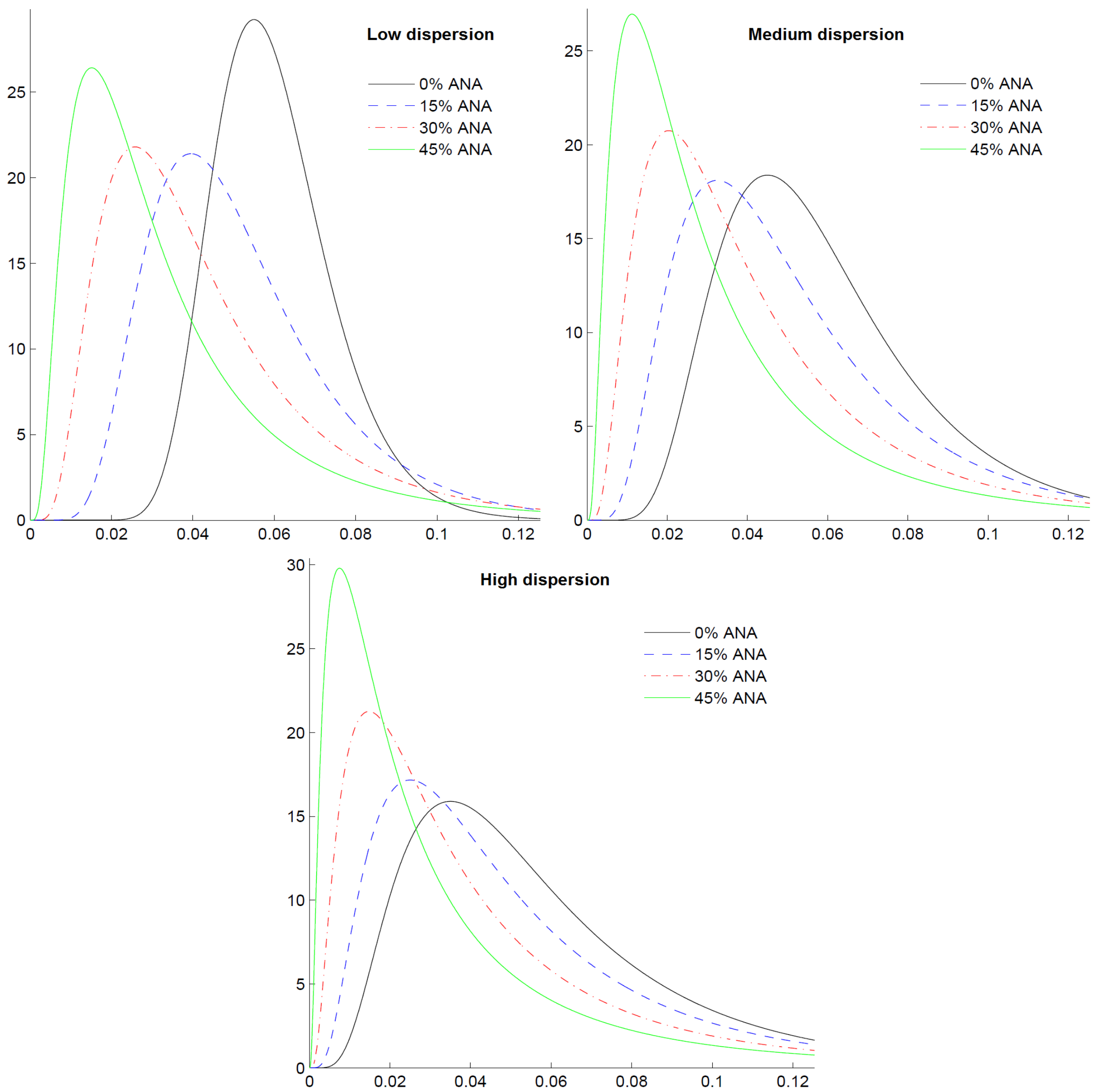

Figure 4.3: Estimated lognormal distributions for parameter two

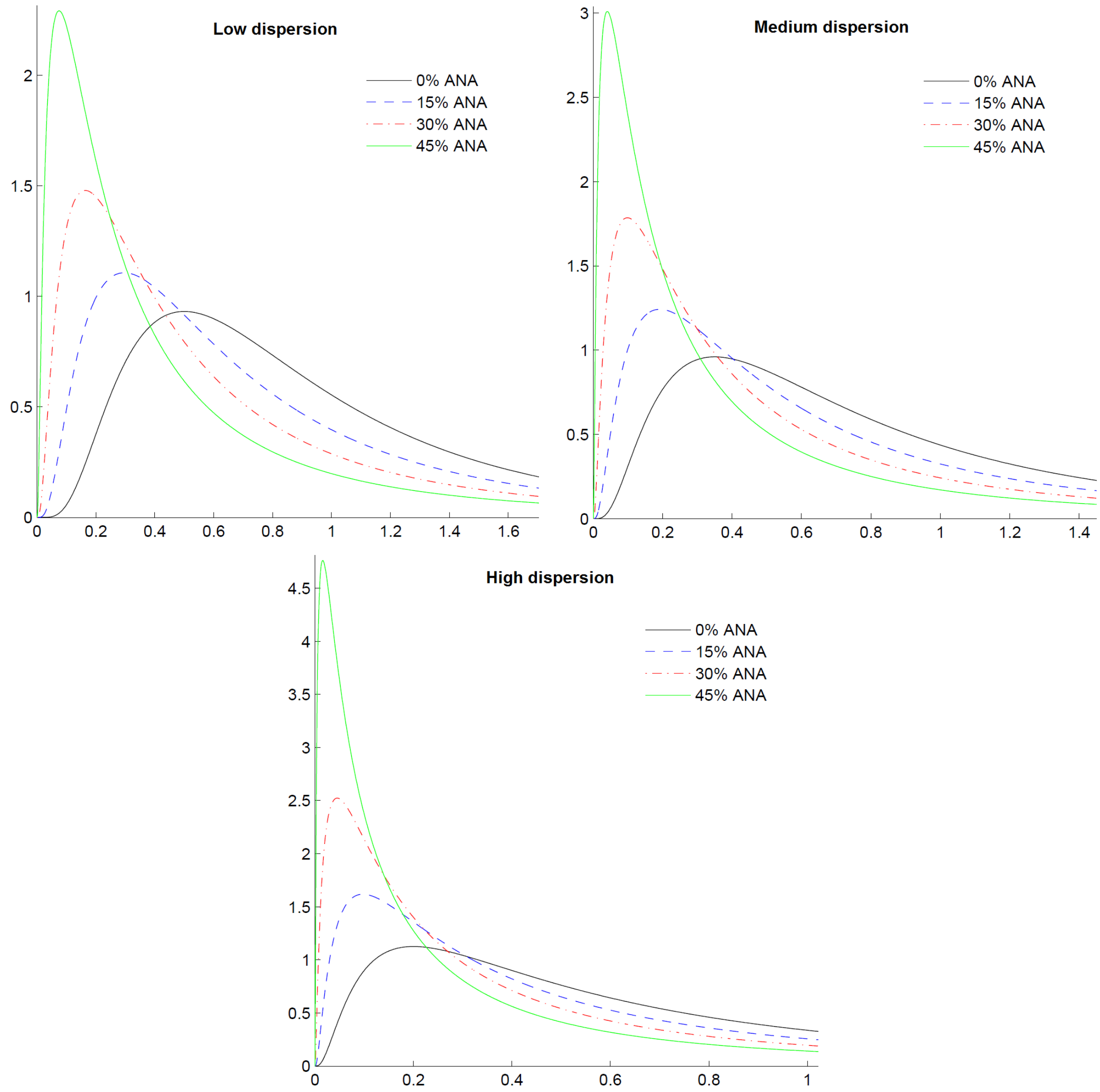

Figure 4.4: Estimated lognormal distributions for parameter three

distribution, a thinning of the tail, and a skewing to the right[8]. The decrease in mode and the skewing to the right are also evidenced as the variance of the true lognormal distribution increases, but ANA heightens these effects. Indeed, as the ANA rate increases, the mode becomes very close to zero, and appears to represent a mass of nonattenders. The thinning of the tail may be deceptive, as it is impractical to clearly plot the distributions over much of the tails, and so the plots do not clearly show what is happening at the extremes. The low dispersion plots for parameters one and two show what might happen: in these cases, the tail not only thins as ANA increases, but also becomes longer. The 99$^{\text{th}}$ percentile will be used below to further consider the impact of ANA on the tail.

Tables 4.11, 4.12 and 4.13 report, for each of the parameters, the true median, mode, standard deviation and ANA rate, the estimated mean, median, mode and standard deviation, and the bias measures that serve as the dependent variable in the regressions[9]. Under full attendance, the true values are well recovered. Attribute nonattendance biases the magnitudes of all three measures of central tendency downwards, with the extent of the bias being proportional to the ANA rate. The bias is most marked for the mode. The true standard deviation has a mixed impact on the estimated mean, but is a further downwards biasing influence on the estimated median and mode. Bias in the standard deviation is somewhat inconsistent, both across dispersion specifications and parameters. Broadly, there is an upward bias in the 99$^{\text{th}}$ percentile, with greater bias for low true dispersion, and as ANA increases.

The SURE model results are contained in Table 4.14. The mean is biased downwards by approximately the same percentage as the ANA rate, but this effect is mitigated increasingly as the true dispersion increases. Conversely, as the true dispersion increases, the downward bias in both the median and mode increases. Overall, ANA has a downward bias on the standard deviation, however, an examination of the residuals suggests that the model fit varies across the three parameters, and so the change in the standard deviation of the lognormal distribution in the presence of ANA might be dependent on the shape of the distribution, in a way that is not apparent. Nonetheless, other measures such as percentiles can provide information that is probably more useful for the lognormal distribution. The regression coefficients estimated in the SURE model support earlier discussion that was informed by inspection of the bias measures. There is an upward bias in the 99$^{\text{th}}$ percentile, which

[8] Assuming a lognormal distribution that has not been transformed to the negative domain.

[9] The bias measures are reported in these tables (where they were not for the other distributions) because the true medians and standard deviations are not round numbers, making the comparison of true values and estimates difficult. The 99$^{\text{th}}$ percentile is only reported as a bias measure, due to space limitations.

Table 4.11: RPL models estimated on datasets with lognormally distributed true sensitivities, parameter one

Tr.	**Actual**[1]				**Estimated**				**Proportion**				
	Med.	**Mode**	**S.D.**	**NA (%)**	**Mean**	**Med.**	**Mode**	**S.D.**	**Mean**	**Med.**	**Mode**	**S.D.**	**99^{th} per.**
1	-0.282	-0.250	0.193	0	-0.296	-0.279	-0.248	0.191	-0.012	-0.010	-0.007	-0.014	-0.021
2	-0.262	-0.200	0.208	0	-0.290	-0.254	-0.196	0.200	-0.035	-0.029	-0.018	-0.041	-0.055
3	-0.238	-0.150	0.229	0	-0.286	-0.229	-0.147	0.217	-0.046	-0.036	-0.017	-0.055	-0.069
4	-0.262	-0.200	0.208	30	-0.212	-0.146	-0.070	0.187	-0.292	-0.441	-0.652	-0.103	0.240
5	-0.238	-0.150	0.229	30	-0.208	-0.129	-0.050	0.204	-0.306	-0.457	-0.668	-0.112	0.083
6	-0.282	-0.250	0.193	30	-0.214	-0.163	-0.094	0.171	-0.286	-0.422	-0.622	-0.117	0.431
7	-0.238	-0.150	0.229	15	-0.247	-0.179	-0.094	0.206	-0.177	-0.249	-0.374	-0.099	-0.001
8	-0.282	-0.250	0.193	15	-0.253	-0.218	-0.162	0.178	-0.156	-0.228	-0.354	-0.077	0.227
9	-0.262	-0.200	0.208	15	-0.250	-0.199	-0.126	0.191	-0.165	-0.240	-0.370	-0.083	0.094
10	-0.282	-0.250	0.193	30	-0.214	-0.164	-0.095	0.170	-0.286	-0.421	-0.619	-0.119	0.424
11	-0.262	-0.200	0.208	30	-0.212	-0.147	-0.070	0.186	-0.294	-0.440	-0.649	-0.109	0.227
12	-0.238	-0.150	0.229	30	-0.211	-0.126	-0.045	0.213	-0.298	-0.470	-0.699	-0.069	0.150
13	-0.238	-0.150	0.229	15	-0.246	-0.179	-0.094	0.206	-0.179	-0.250	-0.375	-0.101	-0.003
14	-0.282	-0.250	0.193	15	-0.253	-0.219	-0.164	0.178	-0.155	-0.224	-0.346	-0.080	0.212
15	-0.262	-0.200	0.208	15	-0.251	-0.200	-0.128	0.191	-0.163	-0.235	-0.362	-0.084	0.089
16	-0.262	-0.200	0.208	45	-0.176	-0.103	-0.035	0.183	-0.412	-0.608	-0.825	-0.120	0.311
17	-0.238	-0.150	0.229	45	-0.174	-0.088	-0.023	0.207	-0.422	-0.630	-0.848	-0.096	0.142
18	-0.282	-0.250	0.193	45	-0.177	-0.117	-0.050	0.163	-0.410	-0.587	-0.798	-0.156	0.542
19	-0.238	-0.150	0.229	45	-0.173	-0.089	-0.024	0.204	-0.424	-0.626	-0.842	-0.112	0.122
20	-0.282	-0.250	0.193	45	-0.176	-0.115	-0.049	0.164	-0.413	-0.594	-0.806	-0.151	0.559
21	-0.262	-0.200	0.208	45	-0.178	-0.101	-0.032	0.191	-0.406	-0.616	-0.839	-0.083	0.375
22	-0.238	-0.150	0.229	45	-0.175	-0.089	-0.023	0.210	-0.416	-0.628	-0.849	-0.083	0.158
23	-0.282	-0.250	0.193	45	-0.178	-0.115	-0.048	0.167	-0.407	-0.593	-0.808	-0.135	0.593
24	-0.262	-0.200	0.208	45	-0.177	-0.102	-0.034	0.187	-0.411	-0.612	-0.832	-0.104	0.339
25	-0.282	-0.250	0.193	0	-0.292	-0.276	-0.245	0.188	-0.026	-0.023	-0.018	-0.029	-0.041
26	-0.262	-0.200	0.208	0	-0.288	-0.254	-0.196	0.199	-0.038	-0.033	-0.021	-0.044	-0.059
27	-0.238	-0.150	0.229	0	-0.285	-0.229	-0.148	0.215	-0.051	-0.039	-0.014	-0.063	-0.081
28	-0.262	-0.200	0.208	0	-0.289	-0.254	-0.196	0.200	-0.037	-0.032	-0.021	-0.042	-0.055
29	-0.238	-0.150	0.229	0	-0.286	-0.230	-0.149	0.216	-0.047	-0.035	-0.010	-0.059	-0.077
30	-0.282	-0.250	0.193	0	-0.292	-0.276	-0.246	0.188	-0.027	-0.023	-0.017	-0.030	-0.044
31	-0.262	-0.200	0.208	30	-0.210	-0.147	-0.072	0.183	-0.298	-0.438	-0.640	-0.123	0.199
32	-0.238	-0.150	0.229	30	-0.208	-0.129	-0.050	0.203	-0.308	-0.459	-0.669	-0.115	0.080
33	-0.282	-0.250	0.193	30	-0.214	-0.163	-0.095	0.170	-0.287	-0.422	-0.621	-0.119	0.426
34	-0.282	-0.250	0.193	15	-0.253	-0.218	-0.162	0.178	-0.156	-0.227	-0.351	-0.079	0.220
35	-0.262	-0.200	0.208	15	-0.250	-0.200	-0.127	0.190	-0.167	-0.238	-0.363	-0.090	0.080
36	-0.238	-0.150	0.229	15	-0.246	-0.179	-0.094	0.206	-0.178	-0.249	-0.374	-0.101	-0.003

1. *mean* =-0.3 for all treatments.

Table 4.12: RPL models estimated on datasets with lognormally distributed true sensitivities, parameter two

Tr.	**Actual[1]**				**Estimated**				**Proportion**				
	Med.	**Mode**	**S.D.**	**NA (%)**	**Mean**	**Med.**	**Mode**	**S.D.**	**Mean**	**Med.**	**Mode**	**S.D.**	**99th per.**
1	-0.058	-0.055	0.037	0	-0.059	-0.057	-0.052	0.038	-0.009	-0.023	-0.051	0.005	0.107
2	-0.055	-0.045	0.040	0	-0.061	-0.056	-0.047	0.041	0.023	0.027	0.033	0.020	0.008
3	-0.050	-0.035	0.044	0	-0.062	-0.052	-0.036	0.045	0.031	0.034	0.042	0.027	0.020
4	-0.050	-0.035	0.044	45	-0.038	-0.022	-0.007	0.040	-0.365	-0.559	-0.787	-0.086	0.238
5	-0.058	-0.055	0.037	45	-0.035	-0.027	-0.015	0.028	-0.410	-0.542	-0.724	-0.240	0.502
6	-0.055	-0.045	0.040	45	-0.036	-0.025	-0.011	0.033	-0.392	-0.546	-0.746	-0.186	0.272
7	-0.050	-0.035	0.044	15	-0.054	-0.042	-0.025	0.043	-0.096	-0.163	-0.282	-0.023	0.099
8	-0.058	-0.055	0.037	15	-0.052	-0.047	-0.040	0.035	-0.132	-0.185	-0.282	-0.076	0.262
9	-0.055	-0.045	0.040	15	-0.053	-0.045	-0.032	0.038	-0.111	-0.172	-0.282	-0.046	0.146
10	-0.055	-0.045	0.040	0	-0.061	-0.056	-0.048	0.040	0.015	0.030	0.059	0.001	-0.046
11	-0.050	-0.035	0.044	0	-0.062	-0.053	-0.038	0.044	0.030	0.047	0.082	0.014	-0.018
12	-0.058	-0.055	0.037	0	-0.060	-0.058	-0.053	0.038	0.005	-0.007	-0.029	0.016	0.100
13	-0.058	-0.055	0.037	30	-0.044	-0.037	-0.026	0.032	-0.270	-0.371	-0.533	-0.153	0.439
14	-0.055	-0.045	0.040	30	-0.044	-0.034	-0.020	0.035	-0.258	-0.370	-0.546	-0.126	0.215
15	-0.050	-0.035	0.044	30	-0.046	-0.031	-0.015	0.041	-0.235	-0.375	-0.583	-0.064	0.183
16	-0.050	-0.035	0.044	45	-0.038	-0.022	-0.008	0.040	-0.360	-0.553	-0.782	-0.084	0.239
17	-0.058	-0.055	0.037	45	-0.036	-0.027	-0.015	0.029	-0.406	-0.542	-0.727	-0.231	0.529
18	-0.055	-0.045	0.040	45	-0.037	-0.025	-0.011	0.034	-0.379	-0.540	-0.747	-0.162	0.315
19	-0.055	-0.045	0.040	0	-0.062	-0.057	-0.048	0.041	0.027	0.039	0.063	0.016	-0.024
20	-0.050	-0.035	0.044	0	-0.062	-0.053	-0.038	0.044	0.033	0.051	0.087	0.015	-0.018
21	-0.058	-0.055	0.037	0	-0.060	-0.058	-0.053	0.038	0.008	-0.006	-0.032	0.021	0.118
22	-0.055	-0.045	0.040	30	-0.045	-0.035	-0.020	0.036	-0.243	-0.364	-0.551	-0.099	0.270
23	-0.050	-0.035	0.044	30	-0.046	-0.032	-0.015	0.041	-0.230	-0.364	-0.566	-0.068	0.168
24	-0.058	-0.055	0.037	30	-0.044	-0.037	-0.026	0.032	-0.264	-0.367	-0.532	-0.145	0.457
25	-0.050	-0.035	0.044	45	-0.038	-0.022	-0.007	0.039	-0.372	-0.565	-0.791	-0.094	0.228
26	-0.058	-0.055	0.037	45	-0.035	-0.027	-0.015	0.028	-0.415	-0.545	-0.724	-0.249	0.479
27	-0.055	-0.045	0.040	45	-0.036	-0.024	-0.011	0.033	-0.394	-0.555	-0.760	-0.174	0.304
28	-0.055	-0.045	0.040	30	-0.045	-0.034	-0.020	0.035	-0.256	-0.369	-0.547	-0.122	0.223
29	-0.050	-0.035	0.044	30	-0.046	-0.031	-0.015	0.041	-0.235	-0.376	-0.585	-0.062	0.188
30	-0.058	-0.055	0.037	30	-0.044	-0.037	-0.026	0.031	-0.275	-0.373	-0.531	-0.161	0.414
31	-0.058	-0.055	0.037	15	-0.052	-0.048	-0.040	0.035	-0.127	-0.180	-0.277	-0.070	0.270
32	-0.055	-0.045	0.040	15	-0.053	-0.045	-0.032	0.038	-0.109	-0.171	-0.281	-0.043	0.151
33	-0.050	-0.035	0.044	15	-0.054	-0.042	-0.025	0.042	-0.098	-0.166	-0.286	-0.024	0.100
34	-0.058	-0.055	0.037	15	-0.052	-0.047	-0.039	0.035	-0.133	-0.187	-0.285	-0.075	0.270
35	-0.055	-0.045	0.040	15	-0.053	-0.045	-0.033	0.038	-0.115	-0.172	-0.275	-0.054	0.126
36	-0.050	-0.035	0.044	15	-0.054	-0.042	-0.025	0.043	-0.096	-0.165	-0.287	-0.022	0.104

1. *mean* =-0.06 for all treatments.

Table 4.13: RPL models estimated on datasets with lognormally distributed true sensitivities, parameter three

Tr.	**Actual**[1]				**Estimated**				**Proportion**				
	Med.	**Mode**	**S.D.**	**NA (%)**	**Mean**	**Med.**	**Mode**	**S.D.**	**Mean**	**Med.**	**Mode**	**S.D.**	**99^{th} per.**
1	-0.705	-0.35	0.861	0	-1.018	-0.722	-0.362	0.872	0.018	0.024	0.035	0.013	0.009
2	-0.585	-0.20	1.037	0	-1.012	-0.568	-0.179	1.094	0.012	-0.028	-0.104	0.054	0.063
3	-0.794	-0.50	0.764	0	-1.007	-0.784	-0.476	0.784	0.007	-0.012	-0.048	0.026	0.052
4	-0.585	-0.20	1.037	30	-0.751	-0.297	-0.047	1.149	-0.249	-0.491	-0.766	0.108	0.083
5	-0.794	-0.50	0.764	30	-0.733	-0.446	-0.166	0.729	-0.267	-0.438	-0.669	-0.045	0.172
6	-0.705	-0.35	0.861	30	-0.736	-0.380	-0.101	0.865	-0.264	-0.461	-0.711	0.006	0.119
7	-0.794	-0.50	0.764	45	-0.588	-0.292	-0.072	0.719	-0.412	-0.632	-0.856	-0.059	0.188
8	-0.705	-0.35	0.861	45	-0.595	-0.242	-0.040	0.888	-0.405	-0.657	-0.886	0.032	0.112
9	-0.585	-0.20	1.037	45	-0.615	-0.180	-0.015	1.273	-0.385	-0.692	-0.923	0.228	0.060
10	-0.705	-0.35	0.861	45	-0.596	-0.241	-0.040	0.893	-0.404	-0.658	-0.887	0.037	0.116
11	-0.585	-0.20	1.037	45	-0.616	-0.179	-0.015	1.288	-0.384	-0.694	-0.925	0.241	0.067
12	-0.794	-0.50	0.764	45	-0.579	-0.298	-0.079	0.682	-0.421	-0.624	-0.842	-0.108	0.127
13	-0.585	-0.20	1.037	15	-0.888	-0.425	-0.097	1.125	-0.112	-0.274	-0.514	0.085	0.100
14	-0.794	-0.50	0.764	15	-0.870	-0.607	-0.296	0.757	-0.130	-0.235	-0.409	-0.010	0.133
15	-0.705	-0.35	0.861	15	-0.879	-0.529	-0.192	0.886	-0.121	-0.249	-0.452	0.030	0.118
16	-0.705	-0.35	0.861	0	-1.002	-0.703	-0.346	0.867	0.002	-0.002	-0.012	0.007	0.010
17	-0.585	-0.20	1.037	0	-1.006	-0.581	-0.193	1.057	0.006	-0.007	-0.033	0.019	0.022
18	-0.794	-0.50	0.764	0	-0.996	-0.790	-0.497	0.761	-0.004	-0.005	-0.006	-0.004	-0.003
19	-0.705	-0.35	0.861	0	-1.005	-0.699	-0.338	0.876	0.005	-0.009	-0.035	0.018	0.028
20	-0.585	-0.20	1.037	0	-1.003	-0.575	-0.189	1.060	0.003	-0.016	-0.053	0.022	0.027
21	-0.794	-0.50	0.764	0	-0.978	-0.784	-0.504	0.740	-0.022	-0.012	0.008	-0.032	-0.046
22	-0.585	-0.20	1.037	45	-0.617	-0.186	-0.017	1.239	-0.383	-0.682	-0.915	0.195	0.047
23	-0.794	-0.50	0.764	45	-0.588	-0.295	-0.074	0.712	-0.412	-0.629	-0.852	-0.068	0.177
24	-0.705	-0.35	0.861	45	-0.600	-0.246	-0.041	0.888	-0.400	-0.651	-0.882	0.032	0.114
25	-0.794	-0.50	0.764	15	-0.875	-0.610	-0.297	0.761	-0.125	-0.231	-0.407	-0.004	0.140
26	-0.705	-0.35	0.861	15	-0.880	-0.528	-0.190	0.891	-0.120	-0.251	-0.458	0.035	0.125
27	-0.585	-0.20	1.037	15	-0.888	-0.424	-0.097	1.128	-0.112	-0.275	-0.517	0.087	0.102
28	-0.794	-0.50	0.764	30	-0.733	-0.444	-0.163	0.733	-0.267	-0.440	-0.674	-0.041	0.180
29	-0.705	-0.35	0.861	30	-0.742	-0.375	-0.096	0.891	-0.258	-0.467	-0.726	0.035	0.152
30	-0.585	-0.20	1.037	30	-0.752	-0.293	-0.045	1.170	-0.248	-0.499	-0.777	0.128	0.097
31	-0.794	-0.50	0.764	15	-0.874	-0.610	-0.297	0.760	-0.126	-0.232	-0.406	-0.006	0.137
32	-0.705	-0.35	0.861	15	-0.880	-0.531	-0.193	0.885	-0.120	-0.247	-0.448	0.028	0.116
33	-0.585	-0.20	1.037	15	-0.881	-0.424	-0.098	1.111	-0.119	-0.275	-0.509	0.071	0.086
34	-0.585	-0.20	1.037	30	-0.750	-0.293	-0.045	1.165	-0.250	-0.499	-0.776	0.123	0.093
35	-0.794	-0.50	0.764	30	-0.732	-0.444	-0.164	0.732	-0.268	-0.440	-0.673	-0.042	0.178
36	-0.705	-0.35	0.861	30	-0.742	-0.377	-0.098	0.884	-0.258	-0.465	-0.721	0.028	0.143

1. *mean* =1 for all treatments.

is accentuated by an increasing ANA rate ('ANA rate' is positive), and mitigated by an increasing true dispersion ('C.V.×ANA' is negative).

Table 4.14: SURE model results - RPL models estimated on datasets with lognormally distributed true sensitivities

	Mean		**Median**		**Mode**		**S.D.**		**99^{th} percentile**	
	Par.	***t*-ratio**	**Par.**	***t*-ratio**	**Par.**	***t*-ratio**	**Par.**	***t*-ratio**	**Par.**	***t*-ratio**
ANA rate	-0.991	-20.38	-0.610	-10.15	-0.418	-3.8	-1.891	-30.59	2.202	12.46
C.V.×ANA	0.132	2.07	-0.860	-11.08	-1.394	-10.14	2.267	28.00	-2.428	-11.01
$\text{Dummy}_{\text{ANA}>0}$			-0.042	-11.93	-0.172	-14.75			0.100	5.29
ρ^2	0.97733		0.98552		0.97616		0.89020		0.74159	

4.3.5 Censored normal distribution

Use of the censored normal distribution in the RPL model (Johnson, 2000; Train and Sonnier, 2005) has the appealing property that it can capture both a point mass at zero, representing ANA, and a distribution of nonzero coefficients, representing preference heterogeneity amongst those who attend to the attribute. Estimation is straightforward in the regular RPL framework, with either classical or Bayesian estimation (Train and Sonnier, 2005). Random parameters logit models with censored normal distributions are the first mechanism evaluated in this chapter that can explicitly handle ANA, and as such their ability to accurately capture ANA is of principal interest. However, as discussed in Section 2.2.2, the censored normal distribution is estimated with only two moments, yet captures three aspects of preference heterogeneity: the ANA rate, a measure of central tendency of the sensitivities of those who attend to the attribute, and a measure of dispersion of these sensitivities. Consequently, one could expect bias in the recovery of these three aspects. This section examines and quantifies any resultant bias, as well as the accuracy of the recovery of ANA.

It must be noted that the true, yet unobservable, distribution of preference heterogeneity in any given empirical context may very closely be approximated by a censored normal distribution. In this situation, a model specified with the censored normal distribution will likely give the best model fit. However, it is highly plausible that in many empirical contexts, the true distribution of sensitivities, including the point mass at zero, cannot be well approximated with the censored normal distribution. This section examines, through simulation, the sorts of biases and inaccuracies that might result when the censored normal distribution is applied to a context in which the true ANA rate and the two moments of a normal distribution vary independently.

The choice of true distribution on which to estimate the censored normal distribution is not straightforward. The normal distribution has the same shape as its censored variant, over the domain of the latter. However, there is a finite probability of sign violation in the normal distribution, and a censoring of these violating coefficients will make it harder to separate the influence of ANA and sign violation on the estimated point mass at zero. The triangular distribution has a different shape, but in its favour is the use, in this chapter, of specifications with no sign violations. Another alternative is generation of the datasets with the censored normal distribution itself. The point mass at zero implied by the true distribution could be combined with the mass resulting from the nonattendance introduced separately, to arrive at a true point mass at zero, to which the estimated point mass at zero could be compared. This last approach is applied herein, as it avoids the complications and confoundment of sign violation, yet utilises a distribution shape consistent with the distribution used for estimation.

A summary of the choice model results is presented across Tables 4.15, 4.16 and 4.17. The true values are reported first, to aid interpretation of the estimated values. The mean and standard deviation reported is conditional on the truncation of the distribution at zero. That is, they describe the continuous portion of the distribution, without being influenced by the point mass at zero. Further details were provided in Section 4.2.1, and the formulas for the conditional mean and variance are in Equation 4.2. The 'censored' column reports the point mass at zero due to the censoring, which is distinct from the additional point mass at zero imposed by the experimental design ('NA'). These two values cannot be merely summed to get a total mass at zero. The coefficients of the distribution are drawn first, and then the ANA rate is applied. Consequently, for the censored normal distribution, two censorings are applied. Firstly, those coefficients in the normal distribution that violate sign are censored to zero. Secondly, irrespective of whether they are already censored, some coefficients of the distribution are censored to zero, at the ANA rate. Some coefficients may be censored twice, and so 'censored' and 'NA' cannot be merely summed. Instead, the final point mass at zero can be calculated as $zero = NA + ((100 - NA)(censored/100))$. The estimated mean and standard deviations, conditional on truncation, are reported in the three tables, together with the percent of coefficients censored by the distribution.

A downward bias in the magnitude of the conditional mean in the presence of ANA can be observed, however, the effect is strongly tempered by the true dispersion. The bias is minimal, slight and pronounced for high, medium and low dispersions specifications, respectively. The bias appears to be nonlinear with respect to the dispersion, with the increase in bias much larger between low and medium than medium and high dispersion specifications. Also, the greater the true ANA rate, the greater the bias. The standard deviation is biased upwards in the presence of ANA, with more upward bias as the true dispersion decreases, and the ANA rate increases. The estimated point masses at zero underestimate the true values in all instances where there is some degree of ANA. The true dispersion influences the accuracy of the retrieved point mass, with low dispersion leading to a much greater underestimation of the bias, when the estimated value is considered as a percentage of the true value. However, the underestimation is tempered by the true ANA rate, and lessened as the ANA rate increases.

The regression and Tobit model results are detailed in Table 4.18. As the true dispersion decreases, '(1/C.V.)×ANA' will increase. The estimated regression coefficient of -0.211 thus implies a downward bias in the mean as the true dispersion decreases. The positive coefficient of 0.181 for 'ANA rate' is more than offset by the negative '(1/C.V.)×ANA' coefficient, as (1/C.V.) evaluates to 1.44, 1.77 and 3.02 for the high, medium and low dispersion specifications, respectively. Under full attendance, the regression model predicts no bias in the mean. Considering the bias in standard deviation, a positive coefficient for '(1/C.V.)×ANA rate' suggests that as either the true dispersion decreases, or the ANA rate increases, the estimated standard deviation will increase. The negative coefficient for the 'ANA rate' main effect has a dampening effect on the extent of the bias, such that only the low dispersion specification results in a large predicted upwards bias. The Tobit model predicting the size of the estimated point mass at zero is illuminating. The marginal effect for the true censored mass is close to one, suggesting that the censored mass implied by the true distribution is accurately recovered. Indeed, this seems to be the case when considering parameter specifications with full attendance, where the true and estimated percentages align well. The extra point mass at zero induced by the ANA rate in the datasets is not fully recovered. Once '$\text{dummy}_{\text{ANA}>0}$' is considered, only an average of 68.5 percent of the true ANA is captured by the censored normal distribution.

Overall, the censored normal distribution performs moderately well, but cannot reliably recover both the true ANA rate, and the sensitivities of those that attend to that attribute. Broadly, there is a downward bias in the mean of the continuous component of the distribution, an upward bias in the standard deviation, and a downward bias in the size of the point mass at zero. The extent of these biases is influenced by the true dispersion and the true ANA rate. As the true dispersion of the preference heterogeneity decreases, the downwards biases in the mean and the size of the point mass at zero become more pronounced. Conversely, high levels of true dispersion lead to more acceptable results, with less bias. Presumably, this is because a censored normal distribution with high dispersion will intrinsically have a lot of mass near zero, and so the introduction of yet more mass at zero will be relatively well

Table 4.15: RPL models estimated on datasets with (censored) normally distributed true sensitivities, parameter one

Tr.	**True**					**Estimated**		
	μ|trunc.	σ|trunc.	censored (%)	NA (%)	zero (%)	μ|trunc.	σ|trunc.	zero (%)
1	-0.3	0.099	0.1	0	0.1	-0.304	0.098	0.2
2	-0.3	0.170	9.4	0	9.4	-0.310	0.171	7.9
3	-0.3	0.208	31.5	0	31.5	-0.306	0.208	27.9
4	-0.3	0.170	9.4	30	36.6	-0.294	0.201	28.7
5	-0.3	0.208	31.5	30	52.1	-0.303	0.224	45.2
6	-0.3	0.099	0.1	30	30.1	-0.260	0.158	14.4
7	-0.3	0.208	31.5	15	41.8	-0.303	0.215	35.7
8	-0.3	0.099	0.1	15	15.1	-0.279	0.141	4.6
9	-0.3	0.170	9.4	15	23.0	-0.300	0.187	17.1
10	-0.3	0.099	0.1	30	30.1	-0.260	0.158	14.4
11	-0.3	0.170	9.4	30	36.6	-0.293	0.200	28.5
12	-0.3	0.208	31.5	30	52.1	-0.304	0.226	45.5
13	-0.3	0.208	31.5	15	41.8	-0.304	0.216	36.0
14	-0.3	0.099	0.1	15	15.1	-0.278	0.139	4.4
15	-0.3	0.170	9.4	15	23.0	-0.300	0.187	17.0
16	-0.3	0.170	9.4	45	50.2	-0.281	0.205	41.4
17	-0.3	0.208	31.5	45	62.3	-0.294	0.228	56.1
18	-0.3	0.099	0.1	45	45.1	-0.241	0.163	27.1
19	-0.3	0.208	31.5	45	62.3	-0.295	0.228	56.2
20	-0.3	0.099	0.1	45	45.1	-0.243	0.165	27.9
21	-0.3	0.170	9.4	45	50.2	-0.279	0.204	41.3
22	-0.3	0.208	31.5	45	62.3	-0.297	0.230	56.7
23	-0.3	0.099	0.1	45	45.1	-0.243	0.165	27.8
24	-0.3	0.170	9.4	45	50.2	-0.280	0.205	41.7
25	-0.3	0.099	0.1	0	0.1	-0.306	0.098	0.1
26	-0.3	0.170	9.4	0	9.4	-0.309	0.170	7.8
27	-0.3	0.208	31.5	0	31.5	-0.306	0.208	27.8
28	-0.3	0.170	9.4	0	9.4	-0.309	0.169	7.7
29	-0.3	0.208	31.5	0	31.5	-0.306	0.208	27.8
30	-0.3	0.099	0.1	0	0.1	-0.306	0.098	0.1
31	-0.3	0.170	9.4	30	36.6	-0.291	0.198	28.1
32	-0.3	0.208	31.5	30	52.1	-0.303	0.225	45.3
33	-0.3	0.099	0.1	30	30.1	-0.261	0.158	14.6
34	-0.3	0.099	0.1	15	15.1	-0.279	0.141	4.6
35	-0.3	0.170	9.4	15	23.0	-0.298	0.186	16.8
36	-0.3	0.208	31.5	15	41.8	-0.304	0.216	35.9

Table 4.16: RPL models estimated on datasets with (censored) normally distributed true sensitivities, parameter two

Tr.	True					Estimated		
	μ|trunc.	σ|trunc.	censored (%)	NA (%)	zero (%)	μ|trunc.	σ|trunc.	zero (%)
1	-0.06	0.020	0.1	0	0.1	-0.059	0.020	0.4
2	-0.06	0.034	9.4	0	9.4	-0.059	0.033	8.6
3	-0.06	0.042	31.5	0	31.5	-0.058	0.040	30.2
4	-0.06	0.042	31.5	45	62.3	-0.054	0.042	59.7
5	-0.06	0.020	0.1	45	45.1	-0.045	0.030	26.1
6	-0.06	0.034	9.4	45	50.2	-0.050	0.036	40.4
7	-0.06	0.042	31.5	15	41.8	-0.056	0.040	39.0
8	-0.06	0.020	0.1	15	15.1	-0.053	0.026	4.2
9	-0.06	0.034	9.4	15	23.0	-0.055	0.035	17.4
10	-0.06	0.034	9.4	0	9.4	-0.059	0.032	7.9
11	-0.06	0.042	31.5	0	31.5	-0.058	0.040	29.7
12	-0.06	0.020	0.1	0	0.1	-0.060	0.019	0.3
13	-0.06	0.020	0.1	30	30.1	-0.049	0.030	14.6
14	-0.06	0.034	9.4	30	36.6	-0.053	0.036	28.3
15	-0.06	0.042	31.5	30	52.1	-0.056	0.042	48.7
16	-0.06	0.042	31.5	45	62.3	-0.054	0.042	59.3
17	-0.06	0.020	0.1	45	45.1	-0.045	0.031	27.0
18	-0.06	0.034	9.4	45	50.2	-0.050	0.037	40.7
19	-0.06	0.034	9.4	0	9.4	-0.059	0.033	8.4
20	-0.06	0.042	31.5	0	31.5	-0.057	0.039	29.1
21	-0.06	0.020	0.1	0	0.1	-0.060	0.019	0.3
22	-0.06	0.034	9.4	30	36.6	-0.054	0.037	29.6
23	-0.06	0.042	31.5	30	52.1	-0.055	0.041	47.7
24	-0.06	0.020	0.1	30	30.1	-0.049	0.029	13.9
25	-0.06	0.042	31.5	45	62.3	-0.054	0.042	59.8
26	-0.06	0.020	0.1	45	45.1	-0.045	0.030	26.7
27	-0.06	0.034	9.4	45	50.2	-0.051	0.037	42.3
28	-0.06	0.034	9.4	30	36.6	-0.053	0.036	29.1
29	-0.06	0.042	31.5	30	52.1	-0.056	0.042	49.5
30	-0.06	0.020	0.1	30	30.1	-0.049	0.029	14.2
31	-0.06	0.020	0.1	15	15.1	-0.053	0.026	4.4
32	-0.06	0.034	9.4	15	23.0	-0.056	0.035	17.3
33	-0.06	0.042	31.5	15	41.8	-0.055	0.040	38.0
34	-0.06	0.020	0.1	15	15.1	-0.053	0.026	4.4
35	-0.06	0.034	9.4	15	23.0	-0.055	0.034	16.9
36	-0.06	0.042	31.5	15	41.8	-0.056	0.040	38.8

Table 4.17: RPL models estimated on datasets with (censored) normally distributed true sensitivities, parameter three

Tr.	True					Estimated		
	μ|trunc.	σ|trunc.	censored (%)	NA (%)	zero (%)	μ|trunc.	σ|trunc.	zero (%)
1	0.9	0.510	9.4	0	9.4	0.892	0.498	8.7
2	0.9	0.624	31.5	0	31.5	0.855	0.594	31.6
3	0.9	0.298	0.1	0	0.1	0.891	0.290	0.2
4	0.9	0.624	31.5	30	52.1	0.807	0.605	47.3
5	0.9	0.298	0.1	30	30.1	0.737	0.439	13.0
6	0.9	0.510	9.4	30	36.6	0.788	0.529	26.2
7	0.9	0.298	0.1	45	45.1	0.682	0.460	26.6
8	0.9	0.510	9.4	45	50.2	0.742	0.537	39.8
9	0.9	0.624	31.5	45	62.3	0.793	0.620	58.8
10	0.9	0.510	9.4	45	50.2	0.742	0.537	39.2
11	0.9	0.624	31.5	45	62.3	0.793	0.620	58.7
12	0.9	0.298	0.1	45	45.1	0.673	0.450	25.5
13	0.9	0.624	31.5	15	41.8	0.818	0.593	40.0
14	0.9	0.298	0.1	15	15.1	0.793	0.394	4.2
15	0.9	0.510	9.4	15	23.0	0.821	0.513	17.2
16	0.9	0.510	9.4	0	9.4	0.876	0.488	8.5
17	0.9	0.624	31.5	0	31.5	0.851	0.589	31.1
18	0.9	0.298	0.1	0	0.1	0.897	0.290	0.2
19	0.9	0.510	9.4	0	9.4	0.873	0.487	8.7
20	0.9	0.624	31.5	0	31.5	0.847	0.586	30.7
21	0.9	0.298	0.1	0	0.1	0.894	0.293	0.2
22	0.9	0.624	31.5	45	62.3	0.787	0.615	58.6
23	0.9	0.298	0.1	45	45.1	0.676	0.454	26.0
24	0.9	0.510	9.4	45	50.2	0.739	0.534	39.3
25	0.9	0.298	0.1	15	15.1	0.790	0.391	4.1
26	0.9	0.510	9.4	15	23.0	0.817	0.508	16.6
27	0.9	0.624	31.5	15	41.8	0.819	0.594	39.9
28	0.9	0.298	0.1	30	30.1	0.732	0.437	13.0
29	0.9	0.510	9.4	30	36.6	0.787	0.531	26.6
30	0.9	0.624	31.5	30	52.1	0.801	0.601	47.7
31	0.9	0.298	0.1	15	15.1	0.796	0.398	4.3
32	0.9	0.510	9.4	15	23.0	0.820	0.512	17.0
33	0.9	0.624	31.5	15	41.8	0.814	0.590	39.6
34	0.9	0.624	31.5	30	52.1	0.797	0.596	47.1
35	0.9	0.298	0.1	30	30.1	0.735	0.439	13.1
36	0.9	0.510	9.4	30	36.6	0.791	0.533	26.7

Table 4.18: SURE and Tobit model results - RPL models estimated on datasets with (censored) normally distributed true sensitivities

	μ		σ		Zero	
	Par.	***t*-ratio**	**Par.**	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	0.181	3.38	-1.515	-16.14	0.735	39.26
(1/C.V.)×ANA	-0.211	-10.85	0.912	26.73		
$\text{Dummy}_{\text{ANA}>0}$	-0.030	-2.65	0.078	3.9	-0.050	-7.92
Censored mass					0.980	73.25
ρ^2	0.74275		0.90135		0.40919	

handled by the distribution, which can accommodate the additional mass with less change to the parameters specifying the distribution. In an empirical context, the true distribution may not be a censored normal, as in these simulations, and the estimation of the censored normal will serve as an approximation to the true distribution. Yet the above results seem to show that as the amount of mass near zero in the true distribution decreases (as distinct from the point mass at zero), the bias in using a censored normal distribution to recover a point mass at zero becomes more acute. Of course the true sensitivities are latent, as is the true ANA rate (with any stated ANA likely only an approximation), and so in an empirical context there is an element of uncertainty as to whether the true values have been recovered with sufficient accuracy. Model fit statistics will provide some clue. However, unless a method that recovers both ANA and attribute sensitivities without bias is available, a model with significant bias might be accepted, as it provides the best model fit with the tools at hand. Consequently, the search continues for techniques that more adequately separate out ANA and preference heterogeneity.

4.4 Independent attribute nonattendance model results

In its most basic form, the IANA model estimates, for each attribute, a single taste coefficient, and a parameter, γ_{c_a}, controlling the probability that the coefficient is zero. The model was introduced in Section 2.1.2, and formalised in Section 3.2. Hole (2011a) interprets the results from this model as if γ_{c_a} is capturing the percentage of the sample that is not attending to an attribute[10]. This body of work argues that such an interpretation may be flawed, for while the nonattendance parameter may capture nonattenders, it may also capture a portion

[10]Precisely, (Hole, 2011a) estimates the percentage that is attending to the attribute. Estimating attendance or nonattendance is equivalent. Hole also introduces covariates into the model that influence that rate of attendance.

of the preference heterogeneity of those who do attend to an attribute. Consequently, the estimated ANA rate might be biased upwards from the true rate. Also, if some of the preference heterogeneity is captured by γ_{c_a}, the magnitude of the mean taste coefficient might be biased upwards, such that it does not capture the mean of the distribution of taste coefficients of those who attend to an attribute. The inability of the IANA model to capture preference heterogeneity (beyond the attendance/nonattendance dichotomy) can readily be acknowledged as a limitation of the model. The extent of bias in the taste coefficients of those interpreted as 'attending' is not so readily apparent. These simulations conveniently allow any bias in the taste coefficients and ANA rates to be quantified, and modelled as a function of the true ANA rate, and normalised measures of dispersion such as the C.V. Four true distributions are tested: the normal, the triangular, the uniform, and the lognormal. In the interests of brevity, the censored normal is not tested.

4.4.1 Normally distributed true distribution

Table 4.19 presents the aggregated results of the IANA choice models estimated on the datasets generated with normally distributed random parameters. The true mean is fixed across treatments for each parameter. The true standard deviations and nonattendance rates (denoted NA in the table for brevity) are reported, together with the estimated means and ANA rates. A clear upwards bias in the magnitude of the estimated sensitivity can be observed, with the extent of bias varying considerably over treatments. The most obvious influence is the true dispersion, with the upwards bias increasing with the magnitude of the true dispersion. The ANA rate appears to have a small influence, with high ANA leading to a greater upward bias, however, the influence is less pronounced than the impact of the true dispersion. The estimated ANA rate is an overestimate of the true ANA rate for all observations. The consistent recovery of nonzero 'ANA' rates for parameter specifications with full attendance supports the argument that the IANA model is confounding nonattendance and preference heterogeneity. The estimated ANA rate increases as the true ANA rate increases, which is promising. Yet the estimated ANA rate also increases as the true measure of dispersion increases, which is problematic.

Table 4.20 presents the regression and Tobit model results. The regression model of bias in the mean exhibits good fit, with a ρ^2 of 0.92700. The ANA rate has a modest upwards contribution to the bias, however, the most prominent contribution is from the C.V. (with a larger coefficient, and larger average sample data points of 0.33, 0.5 and 1, compared to 0,

Table 4.19: IANA models estimated on datasets with normally distributed true sensitivities

Tr.	**Parameter One**				**Parameter Two**				**Parameter Three**			
	Actual		**Estimated**		**Actual**		**Estimated**		**Actual**		**Estimated**	
	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)
1	0.1	0	-0.319	8.4	0.02	0	-0.065	9.1	0.6	0	1.223	29.7
2	0.2	0	-0.389	27.1	0.04	0	-0.083	35.5	0.9	0	1.538	46.0
3	0.3	0	-0.460	37.0	0.06	0	-0.107	50.8	0.3	0	0.918	7.5
4	0.2	30	-0.391	48.3	0.06	45	-0.109	74.0	0.9	30	1.537	62.7
5	0.3	30	-0.472	55.9	0.02	45	-0.068	54.8	0.3	30	0.964	38.2
6	0.1	30	-0.329	37.2	0.04	45	-0.088	66.7	0.6	30	1.229	52.5
7	0.3	15	-0.461	46.4	0.06	15	-0.109	58.6	0.3	45	0.949	51.5
8	0.1	15	-0.320	22.0	0.02	15	-0.065	25.2	0.6	45	1.252	64.7
9	0.2	15	-0.384	36.8	0.04	15	-0.084	45.8	0.9	45	1.591	73.1
10	0.1	30	-0.328	36.9	0.04	0	-0.082	31.9	0.6	45	1.251	64.6
11	0.2	30	-0.393	48.4	0.06	0	-0.106	48.7	0.9	45	1.572	72.5
12	0.3	30	-0.473	55.9	0.02	0	-0.063	9.2	0.3	45	0.970	52.7
13	0.3	15	-0.467	47.7	0.02	30	-0.067	42.9	0.9	15	1.515	53.6
14	0.1	15	-0.321	21.8	0.04	30	-0.087	55.9	0.3	15	0.969	23.3
15	0.2	15	-0.387	37.4	0.06	30	-0.113	67.2	0.6	15	1.204	41.0
16	0.2	45	-0.402	61.5	0.06	45	-0.115	75.2	0.6	0	1.175	28.8
17	0.3	45	-0.484	67.7	0.02	45	-0.067	54.7	0.9	0	1.484	43.7
18	0.1	45	-0.329	51.1	0.04	45	-0.088	66.7	0.3	0	0.942	7.2
19	0.3	45	-0.489	67.8	0.04	0	-0.082	33.9	0.6	0	1.182	29.5
20	0.1	45	-0.333	52.4	0.06	0	-0.107	48.7	0.9	0	1.513	43.9
21	0.2	45	-0.402	61.7	0.02	0	-0.063	9.7	0.3	0	0.932	7.4
22	0.3	45	-0.482	66.8	0.04	30	-0.085	55.4	0.9	45	1.552	72.1
23	0.1	45	-0.326	50.1	0.06	30	-0.110	65.1	0.3	45	0.969	51.5
24	0.2	45	-0.401	60.8	0.02	30	-0.067	40.9	0.6	45	1.242	64.6
25	0.1	0	-0.316	6.2	0.06	45	-0.113	74.6	0.3	15	0.975	24.0
26	0.2	0	-0.387	25.8	0.02	45	-0.070	57.5	0.6	15	1.229	42.3
27	0.3	0	-0.465	37.8	0.04	45	-0.089	68.3	0.9	15	1.512	53.7
28	0.2	0	-0.383	24.6	0.04	30	-0.089	58.0	0.3	30	0.984	39.6
29	0.3	0	-0.463	37.1	0.06	30	-0.113	67.6	0.6	30	1.202	52.6
30	0.1	0	-0.315	7.5	0.02	30	-0.066	40.3	0.9	30	1.579	63.8
31	0.2	30	-0.395	48.8	0.02	15	-0.065	26.2	0.3	15	0.961	23.8
32	0.3	30	-0.476	56.7	0.04	15	-0.084	46.3	0.6	15	1.205	41.3
33	0.1	30	-0.331	38.2	0.06	15	-0.107	57.7	0.9	15	1.524	52.5
34	0.1	15	-0.322	23.0	0.02	15	-0.064	24.6	0.9	30	1.555	62.7
35	0.2	15	-0.386	36.9	0.04	15	-0.084	46.0	0.3	30	0.974	39.1
36	0.3	15	-0.465	47.0	0.06	15	-0.109	59.1	0.6	30	1.204	52.6

0.15, 0.3 and 0.45 for ANA). The Tobit model has the estimated ANA rate, represented as a proportion, as the dependent variable. Considering the marginal effects, both the true ANA rate and the C.V. have an upwards influence on the estimated ANA rate. The probabilities that each observation have not been censored evaluate to 99.89 percent or higher for all observations. Both of these models are consistent with the inspection of the data. Assuming that the true attribute sensitivity is normally distributed, the IANA model induces a clear upwards bias in both the point estimate of this sensitivity (relative to the mean of the true distribution), and the estimated ANA rate.

Table 4.20: Regression and Tobit model results - IANA models estimated on datasets with normally distributed true sensitivities

	μ		**ANA**	
	Par.	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	0.155	3.86	0.802	35.33
C.V.	0.936	33.68	0.406	45.86
Constant	-0.277	-16.88		
ρ^2	0.92700		0.46419	

4.4.2 Triangularly distributed true distribution

Table 4.21 presents the aggregated results of the IANA choice models estimated on the datasets generated with triangularly distributed random parameters. Whilst an upwards bias of the mean sensitivity can again be observed, the magnitude of the bias is considerably less than the normal distribution. The true ANA rate and true dispersion both have a positive influence on the two measures of interest: the bias of the mean and the estimated ANA rate. As the true dispersion decreases, both the estimated mean and estimated ANA rate approach their true values. This is plausible and encouraging. The IANA model is constructed with the implicit assumption (not acknowledged in the literature) that those decision makers who attend to an attribute exhibit no preference heterogeneity. So long as this assumption holds, the model performs well and retrieves the true sensitivities and ANA rates. However, the results also clearly show that if preference heterogeneity is present, upwards bias is introduced into the magnitudes of both measures. Interestingly, the estimated values for datasets generated with the normal distribution, presented in the previous section, did not collapse back to the true values as the true dispersion tended to zero. It is possible that the unbounded nature of the normal distribution leads to confoundment between ANA and preference heterogeneity, even where the true dispersion is quite low.

Table 4.21: IANA models estimated on datasets with triangularly distributed true sensitivities

Tr.	**Parameter One**				**Parameter Two**				**Parameter Three**			
	Actual		**Estimated**		**Actual**		**Estimated**		**Actual**		**Estimated**	
	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)
1	0.1	0	-0.302	0.7	0.02	0	-0.061	2.0	0.6	0	0.940	4.8
2	0.2	0	-0.310	4.3	0.04	0	-0.062	6.4	0.9	0	0.997	12.6
3	0.3	0	-0.328	10.6	0.06	0	-0.067	14.9	0.3	0	0.894	1.6
4	0.2	30	-0.320	34.8	0.06	45	-0.070	55.8	0.9	30	1.023	41.1
5	0.3	30	-0.341	39.2	0.02	45	-0.062	47.4	0.3	30	0.907	32.2
6	0.1	30	-0.307	31.3	0.04	45	-0.064	49.7	0.6	30	0.947	34.8
7	0.3	15	-0.332	25.0	0.06	15	-0.068	29.6	0.3	45	0.906	47.2
8	0.1	15	-0.305	16.2	0.02	15	-0.061	16.7	0.6	45	0.945	49.1
9	0.2	15	-0.314	19.7	0.04	15	-0.064	22.6	0.9	45	1.028	54.8
10	0.1	30	-0.307	31.3	0.04	0	-0.063	6.0	0.6	45	0.948	49.2
11	0.2	30	-0.320	34.8	0.06	0	-0.067	14.1	0.9	45	1.029	54.9
12	0.3	30	-0.341	39.3	0.02	0	-0.061	2.8	0.3	45	0.905	46.9
13	0.3	15	-0.335	25.9	0.02	30	-0.062	34.4	0.9	15	1.010	27.3
14	0.1	15	-0.305	16.1	0.04	30	-0.064	35.6	0.3	15	0.907	16.0
15	0.2	15	-0.314	19.7	0.06	30	-0.070	43.2	0.6	15	0.940	20.3
16	0.2	45	-0.320	49.3	0.06	45	-0.070	55.5	0.6	0	0.920	4.3
17	0.3	45	-0.346	54.2	0.02	45	-0.062	48.0	0.9	0	0.981	11.8
18	0.1	45	-0.306	45.8	0.04	45	-0.063	48.6	0.3	0	0.902	1.0
19	0.3	45	-0.346	54.3	0.04	0	-0.062	6.2	0.6	0	0.914	4.3
20	0.1	45	-0.307	46.7	0.06	0	-0.067	13.4	0.9	0	0.994	11.7
21	0.2	45	-0.321	49.2	0.02	0	-0.060	2.2	0.3	0	0.899	1.0
22	0.3	45	-0.343	53.2	0.04	30	-0.065	37.3	0.9	45	1.033	55.3
23	0.1	45	-0.306	45.6	0.06	30	-0.069	41.7	0.3	45	0.910	46.4
24	0.2	45	-0.320	48.6	0.02	30	-0.061	32.3	0.6	45	0.949	49.7
25	0.1	0	-0.303	0.5	0.06	45	-0.070	54.7	0.3	15	0.906	16.3
26	0.2	0	-0.310	3.6	0.02	45	-0.061	47.1	0.6	15	0.949	21.0
27	0.3	0	-0.329	10.9	0.04	45	-0.065	52.4	0.9	15	1.023	28.4
28	0.2	0	-0.309	3.3	0.04	30	-0.064	37.1	0.3	30	0.912	32.1
29	0.3	0	-0.328	10.2	0.06	30	-0.070	44.4	0.6	30	0.956	36.6
30	0.1	0	-0.302	0.7	0.02	30	-0.061	32.0	0.9	30	1.031	41.3
31	0.2	30	-0.319	34.3	0.02	15	-0.060	17.0	0.3	15	0.907	16.6
32	0.3	30	-0.341	39.7	0.04	15	-0.063	22.4	0.6	15	0.939	20.9
33	0.1	30	-0.307	31.7	0.06	15	-0.068	28.2	0.9	15	1.013	26.4
34	0.1	15	-0.305	16.6	0.02	15	-0.061	17.3	0.9	30	1.020	40.4
35	0.2	15	-0.314	19.3	0.04	15	-0.063	21.9	0.3	30	0.909	31.7
36	0.3	15	-0.333	25.3	0.06	15	-0.068	29.8	0.6	30	0.945	35.9

The regression and Tobit model results for the triangular distribution are contained in Table 4.22. Model fits are good, and the marginal effects are consistent with the above discussion. In the Tobit model of estimated ANA, the marginal effect for the true ANA rate is close to one. Together with a small constant of -0.032, this implies that if there is no true dispersion (i.e., the C.V. is zero), the estimated ANA rate is almost equal to the true ANA rate. As discussed, true dispersion biases both the estimated mean and ANA rate upwards, as evidenced by the positive marginal effects for 'C.V'. Only the observations with full attendance and low dispersion have a probability of not being censored of less than 100 percent, with a probability of 96.53 percent.

Table 4.22: Regression and Tobit model results - IANA models estimated on datasets with triangularly distributed true sensitivities

	μ		**ANA**	
	Par.	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	0.072	7.76	0.983	119.44
C.V.	0.178	36.27	0.145	28.58
Constant	-0.069	-16.13	-0.032	-7.75
ρ^2	0.91748		0.42476	

4.4.3 Uniformly distributed true distribution

Table 4.23 presents the aggregated results of the IANA choice models estimated on the datasets generated with uniformly distributed random parameters. The pattern is very similar to those observed with the triangular distribution, with true dispersion and true ANA biasing the magnitude of the estimated mean and the estimated ANA upwards. Again these two estimated measures approach their true values as true preference heterogeneity narrows around the mean. The key difference to the triangular distribution is a larger bias in both measures with the uniform distribution, for any given treatment. This is likely due to the uniform distribution having more mass closer to zero then the triangular distribution when the mean and spread are the same, as discussed previously in Section 4.3.3. This mass in turn can be approximated by the nonattendance parameter in the IANA model, thus leading to greater confoundment between nonattendance and preference heterogeneity.

The regression and Tobit model results are detailed in Table 4.24. Both models bear a distinct resemblance to their equivalents estimated on the triangular datasets. The marginal effects of 'ANA' and 'C.V.' are greater for the bias in the estimated mean, and greater for 'C.V'. for the estimated rate of nonattendance. Again, only the observations with full

Table 4.23: IANA models estimated on datasets with uniformly distributed true sensitivities

Tr.	**Parameter One**				**Parameter Two**				**Parameter Three**			
	Actual		**Estimated**		**Actual**		**Estimated**		**Actual**		**Estimated**	
	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)	σ	**NA** (%)	μ	**NA** (%)
1	0.1	0	-0.303	1.6	0.02	0	-0.061	3.2	0.6	0	0.995	11.2
2	0.2	0	-0.327	11.0	0.04	0	-0.066	13.7	0.9	0	1.146	26.8
3	0.3	0	-0.375	24.3	0.06	0	-0.077	29.0	0.3	0	0.882	2.4
4	0.2	30	-0.341	39.9	0.06	45	-0.083	64.2	0.9	30	1.165	50.3
5	0.3	30	-0.388	47.9	0.02	45	-0.063	49.9	0.3	30	0.911	34.0
6	0.1	30	-0.312	32.8	0.04	45	-0.069	54.3	0.6	30	1.005	39.8
7	0.3	15	-0.377	35.8	0.06	15	-0.079	41.4	0.3	45	0.919	49.4
8	0.1	15	-0.308	17.6	0.02	15	-0.062	19.2	0.6	45	1.011	53.7
9	0.2	15	-0.330	25.1	0.04	15	-0.067	28.6	0.9	45	1.186	63.0
10	0.1	30	-0.312	32.7	0.04	0	-0.066	12.3	0.6	45	1.014	53.9
11	0.2	30	-0.340	39.7	0.06	0	-0.078	27.6	0.9	45	1.186	62.9
12	0.3	30	-0.389	48.0	0.02	0	-0.060	3.7	0.3	45	0.913	48.8
13	0.3	15	-0.382	37.0	0.02	30	-0.063	37.4	0.9	15	1.150	38.4
14	0.1	15	-0.307	17.3	0.04	30	-0.068	41.0	0.3	15	0.914	17.4
15	0.2	15	-0.331	25.5	0.06	30	-0.082	53.2	0.6	15	0.994	26.5
16	0.2	45	-0.342	54.2	0.06	45	-0.082	64.1	0.6	0	0.963	10.5
17	0.3	45	-0.399	61.7	0.02	45	-0.064	51.2	0.9	0	1.120	25.3
18	0.1	45	-0.310	47.2	0.04	45	-0.068	53.6	0.3	0	0.904	1.9
19	0.3	45	-0.397	61.4	0.04	0	-0.065	12.8	0.6	0	0.960	10.8
20	0.1	45	-0.313	48.5	0.06	0	-0.077	26.9	0.9	0	1.138	25.3
21	0.2	45	-0.343	53.9	0.02	0	-0.061	3.4	0.3	0	0.896	1.9
22	0.3	45	-0.395	60.6	0.04	30	-0.069	43.1	0.9	45	1.181	63.0
23	0.1	45	-0.310	46.7	0.06	30	-0.080	50.9	0.3	45	0.919	47.6
24	0.2	45	-0.341	53.1	0.02	30	-0.063	34.8	0.6	45	1.015	54.6
25	0.1	0	-0.303	1.2	0.06	45	-0.080	62.5	0.3	15	0.917	18.0
26	0.2	0	-0.325	9.8	0.02	45	-0.063	50.1	0.6	15	1.007	27.2
27	0.3	0	-0.377	24.3	0.04	45	-0.072	58.9	0.9	15	1.161	39.3
28	0.2	0	-0.323	9.1	0.04	30	-0.069	42.8	0.3	30	0.926	34.2
29	0.3	0	-0.374	23.4	0.06	30	-0.083	54.8	0.6	30	1.014	42.2
30	0.1	0	-0.302	1.6	0.02	30	-0.062	34.2	0.9	30	1.192	51.2
31	0.2	30	-0.339	39.4	0.02	15	-0.061	19.1	0.3	15	0.914	18.3
32	0.3	30	-0.390	48.5	0.04	15	-0.067	29.1	0.6	15	0.989	26.7
33	0.1	30	-0.313	33.5	0.06	15	-0.078	39.8	0.9	15	1.154	37.3
34	0.1	15	-0.308	18.1	0.02	15	-0.062	19.0	0.9	30	1.175	50.0
35	0.2	15	-0.330	24.9	0.04	15	-0.067	28.1	0.3	30	0.917	33.6
36	0.3	15	-0.380	36.4	0.06	15	-0.079	41.8	0.6	30	1.001	41.4

attendance and low dispersion have a probability of not being censored of less than 100 percent, with a probability of 87.63 percent.

Table 4.24: Regression and Tobit model results - IANA models estimated on datasets with uniformly distributed true sensitivities

	μ		**ANA**	
	Par.	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	0.118	7.15	0.936	70.18
C.V.	0.416	47.16	0.278	33.8
Constant	-0.156	-20.42	-0.050	-7.59
ρ^2	0.94285		0.44264	

4.4.4 Lognormally distributed true distribution

Table 4.25 presents the aggregated results of the IANA choice models estimated on the datasets generated with lognormally distributed random parameters. As the standard deviation of the true distribution increases, the magnitude of the estimated sensitivity and ANA increases. An increase in the true ANA rate leads to a higher estimated ANA rate, but the impact on the sensitivity is not strongly pronounced. Any comparison of the estimated sensitivity with the true lognormal distribution is hampered by the asymmetry of the lognormal distribution used to generate the datasets. For this analysis, the point estimate of the sensitivity is compared to the mean of the lognormal distribution, which remains the same across each dispersion specification for each parameter. An examination of the estimated ANA rate is straightforward, as while the true distribution has some coefficient close to zero, it has no mass at zero, beyond what is induced by the ANA. Clearly, the estimated ANA rate is not accurate, and overestimates the true rate in all cases. Even when the true ANA rate is zero, the estimated ANA rate is greater than zero, and as high as 55.5 percent in one instance.

Table 4.26 presents the regression and Tobit model results. The model fit for the bias in the mean is lower than for the other distributions, perhaps reflecting the difficulty in comparing a point estimate to the asymmetry of the lognormal distribution. Nonetheless, as with other distributions, the ANA rate and true dispersion both have an upwards bias on the magnitude of the sensitivity, in line with the above inspection of the data. The marginal effects of the Tobit model are also consistent with the above observations. Also, as with previous distributions, as the true dispersion approaches zero, the estimated ANA rate approaches the true rate (since with 'C.V.'=0, the total marginal effect will be $0.853+0.214 = 1.067$).

Table 4.25: IANA models estimated on datasets with lognormally distributed true sensitivities

Tr.	Parameter One				Parameter Two				Parameter Three			
	Actual[1]		Estimated		Actual[2]		Estimated		Actual[3]		Estimated	
	σ	NA (%)	μ	NA (%)	σ	NA (%)	μ	NA (%)	σ	NA (%)	μ	NA (%)
1	low	0	-0.319	14.6	low	0	-0.061	8.2	med.	0	1.385	40.1
2	med.	0	-0.334	24.5	med.	0	-0.073	24.1	high	0	1.608	55.5
3	high	0	-0.346	31.0	high	0	-0.084	35.3	low	0	1.270	32.5
4	med.	30	-0.334	47.9	high	45	-0.099	69.4	high	30	1.654	71.0
5	high	30	-0.343	52.3	low	45	-0.071	55.7	low	30	1.309	55.3
6	low	30	-0.318	41.0	med.	45	-0.082	61.0	med.	30	1.423	61.3
7	high	15	-0.337	39.3	high	15	-0.086	44.3	low	45	1.398	68.9
8	low	15	-0.310	25.2	low	15	-0.065	23.8	med.	45	1.529	73.4
9	med.	15	-0.321	32.8	med.	15	-0.075	34.9	high	45	1.810	80.7
10	low	30	-0.312	39.8	med.	0	-0.069	15.8	med.	45	1.526	73.4
11	med.	30	-0.327	47.0	high	0	-0.078	28.1	high	45	1.820	80.9
12	high	30	-0.340	52.0	low	0	-0.063	7.3	low	45	1.375	68.3
13	high	15	-0.351	43.1	low	30	-0.071	45.0	high	15	1.638	63.6
14	low	15	-0.318	27.0	med.	30	-0.078	48.4	low	15	1.281	42.2
15	med.	15	-0.334	35.3	high	30	-0.094	59.3	med.	15	1.421	52.3
16	med.	45	-0.349	61.0	high	45	-0.101	69.9	med.	0	1.330	38.5
17	high	45	-0.372	66.2	low	45	-0.075	59.3	high	0	1.529	52.4
18	low	45	-0.325	54.9	med.	45	-0.083	61.6	low	0	1.206	26.7
19	high	45	-0.370	66.3	med.	0	-0.073	21.7	med.	0	1.353	40.5
20	low	45	-0.337	57.7	high	0	-0.082	31.9	high	0	1.538	52.7
21	med.	45	-0.342	60.1	low	0	-0.064	8.5	low	0	1.223	28.5
22	high	45	-0.356	64.1	med.	30	-0.077	48.1	high	45	1.764	79.9
23	low	45	-0.319	53.1	high	30	-0.087	53.6	low	45	1.331	66.3
24	med.	45	-0.335	58.3	low	30	-0.067	39.0	med.	45	1.491	72.5
25	low	0	-0.311	12.3	high	45	-0.096	68.8	low	15	1.296	43.5
26	med.	0	-0.327	21.3	low	45	-0.073	57.8	med.	15	1.446	53.2
27	high	0	-0.345	31.3	med.	45	-0.090	67.5	high	15	1.662	64.3
28	med.	0	-0.319	18.5	med.	30	-0.079	50.0	low	30	1.362	57.2
29	high	0	-0.337	28.6	high	30	-0.095	60.5	med.	30	1.536	65.7
30	low	0	-0.307	12.0	low	30	-0.068	41.8	high	30	1.722	72.7
31	med.	30	-0.332	47.0	low	15	-0.067	25.5	low	15	1.280	42.6
32	high	30	-0.354	54.6	med.	15	-0.077	37.5	med.	15	1.416	52.5
33	low	30	-0.326	43.7	high	15	-0.087	45.3	high	15	1.581	61.8
34	low	15	-0.314	27.3	low	15	-0.066	25.2	high	30	1.670	71.4
35	med.	15	-0.325	32.6	med.	15	-0.075	34.4	low	30	1.334	55.9
36	high	15	-0.344	41.4	high	15	-0.087	46.1	med.	30	1.500	64.5

1. $mean_{lognormal}$ =-0.3 for all treatments.
2. $mean_{lognormal}$ =-0.06 for all treatments.
3. $mean_{lognormal}$ =1.0 for all treatments.

Table 4.26: Regression and Tobit model results - IANA models estimated on datasets with lognormally distributed true sensitivities

	Mean		**ANA**	
	Par.	***t*-ratio**	**Marginal effects**	***t*-ratio**
ANA rate	0.336	3.34	0.853	17.81
C.V.	0.093	6.11	0.064	8.40
Constant	0.128	4.87	0.214	13.90
ρ^2	0.30276		0.46619	

4.5 Discussion

The results in this chapter have demonstrated the potentially severe impact of ANA on the accuracy of the mean and dispersion of randomly distributed parameters. Further, they have allowed the link between ANA and sign violation to be probed, and the mechanics of how RPL models cope with ANA to be explored. Table 4.27 provides a summary of the findings for all of the RPL models estimated.

Table 4.27: Summary of findings for RPL models

Distribution Measure	**Dominant changes in measure**
Normal	
Mean	Decreases as ANA increases, by roughly same percentage as ANA
Standard deviation	Mostly increases in presence of ANA, especially when true value is low
Sign violation %	Increases as ANA increases, at about 1/3 of ANA rate
Triangular	
Mean	Decreases as ANA increases, by roughly same percentage as ANA
Spread	Increases in presence of ANA, more so as true spread decreases
Sign violation	Increases as ANA increases and as true spread increases
Uniform	
Mean	Decreases as ANA increases, by roughly same percentage as ANA
Spread	Increases in presence of ANA, more so as true spread decreases
Sign violation	Increases as ANA increases and as true spread increases
Lognormal	
Mean	Decreases as ANA increases, by roughly same percentage as ANA
Median	Decreases as ANA increases, especially for high true dispersion
Mode	Decreases as ANA increases, especially for high true dispersion
Standard deviation	Decreases as ANA increases, but not a strong finding
99th percentile	Increases as ANA increases, but decreases as true dispersion increases
Censored normal	
Mean	Decreases as ANA increases, more so as true standard deviation decreases
Standard deviation	Increases in presence of ANA, more so as true standard deviation decreases
Zero mass	Increases as ANA increases, but ANA rate not fully recovered

The impact of ANA on the mean of the RP distributions is unsurprising, and consistent across the distributions tested. The mean is biased downwards by approximately the same percentage as the ANA rate. This is consistent with, but extends upon Hoyos et al. (2010), who made the same finding when using fixed coefficients in their simulated datasets.

The impact on the measure of dispersion for the normal, triangular and uniform distributions is more subtle. Attribute nonattendance has an upwards biasing influence on the standard deviation or spread, as with Mariel et al. (2011), however, the extent of the bias is accentuated for lower standard deviations or spreads. This is plausible, for under higher true levels of dispersion, there will be more mass close to zero, which can serve to approximate ANA. When small to start with, the measure of dispersion must be increased in magnitude relatively more to place the requisite mass near zero and approximate ANA. The shape of the distribution also plays a role. The uniform distribution exhibited less downward bias in the spread than the triangular distribution, ceteris paribus, for it has more mass near the limits of its domain than the triangular distribution.

The implication of ANA biasing the measure of dispersion is that some of what might readily be interpreted as preference heterogeneity might actually be ANA[11]. This is a finding that has been observed empirically in a number of studies (Campbell et al., 2012; Hess et al., 2011). By using known ANA rates, this chapter quantifies the extent of this bias. It has been shown that the bias is particularly prominent if there is little preference heterogeneity. At the extreme, there may be no preference heterogeneity at all, just a single sensitivity and nonattendance, yet this true state may be misrepresented by the estimation of a significant measure of dispersion.

Finally, a clear causal link has been demonstrated between ANA and sign violation, supporting the finding of Hensher (2007). The rate of sign violation increases not only as ANA increases, as expected, but also as the true measure of dispersion increases. That is, the more mass near zero before ANA is introduced, the more there will be sign violation after it is introduced. Thus, true distributions with some low sensitivities are more prone to sign violation in the presence of ANA.

Consider now the evaluation of the analytical techniques for handling ANA. Whilst the censored normal distribution performs moderately well, it cannot accurately recover the true parameters and ANA rates. Broadly, there is a downward bias in the mean of the continuous component of the distribution, an upward bias in the standard deviation, and a downward

[11]ANA could be considered as an extreme form of preference heterogeneity, but it is sufficiently different in nature to draw a distinction between the two.

bias in the size of the point mass at zero. In the dispersion specifications with moderate and large standard deviations, some ANA was induced through the distribution itself. So long as no further ANA was introduced, the distribution was accurately recovered, and the correct ANA inferred. However, once further ANA was added, neither the ANA, nor the continuous component of utility could be accurately estimated. Thus, the censored normal works well if the ANA rate induced by the censoring is consistent with the true specification of the rest of the distribution. If the ANA rate varies independently, the accuracy is compromised, since only two structural parameters are estimated to represent three moments - the mean, the standard deviation, and the size of the ANA point mass at zero. Overall, the censored normal will work well in a limited set of circumstances, but not well handle a wider range of ANA rates and distributions of preference heterogeneity.

The simulations have allowed the performance of the IANA model to be tested, and the results are not encouraging. Whilst the ANA rate and mean sensitivities are recovered with increasing accuracy as the extent of the true preference heterogeneity diminishes, bias is observed in the presence of heterogeneity. Specifically, the IANA mode is prone to overestimating both the mean of the sensitivities, and the ANA rate. Given that preference heterogeneity can be expected for a wide range of attributes across many choice contexts, such bias is of great concern, and the IANA model may not be appropriate in many circumstances.

To reiterate, the bias in the recovery of the mean sensitivity is the concern. The IANA model can be criticised for only estimating a single nonzero sensitivity per attribute, given the ubiquity of RPL and LC models that can capture more than a single point estimate. Whilst a limitation of the model, a point estimate is not inherently problematic, so long as it is meaningful in the context of true preference heterogeneity, for example by representing the mean sensitivity. The mean is not being recovered in the simulations, which is of concern. That the scope of the IANA model is limited with respect to preference heterogeneity is not itself a problem. What is a problem is that the accuracy of the IANA model is compromised by the preference heterogeneity it does not seek to accommodate.

Clearly the bias in the IANA model is undesirable. Overestimates of mean sensitivities will have a biasing influence on WTP. Overestimates of the ANA rate will have a distinct impact on WTP, as ANA will either result in a zero WTP for non-cost attributes, or an 'infinite' WTP for cost. The danger is that by trying to avoid the biasing influence of ANA by analytically identifying and accommodating ANA, further bias is introduced. The shortcoming of the IANA model is its inability to capture preference heterogeneity, beyond the mean sensitivity/ANA dichotomy. If more than a single sensitivity could be estimated, then the confounding between ANA and low sensitivities could be reduced. This was demonstrated by Campbell et al. (2012), who in addition to a fixed point mass at zero, estimated multiple freely estimable point masses in an empirical setting. As more point masses were introduced, the ANA rate decreased. With the RPANA model, this thesis proposes an alternative, whereby preference heterogeneity is captured parsimoniously through a RP distribution. The next chapter will test the performance of the RPANA model on the simulated datasets introduced in this chapter, with particular attention paid to whether any bias is introduced into the model outputs.

Note that in all of the simulations in this chapter, ANA is treated as exogenous. Attribute nonattendance is assumed to vary independently of any properties of the design, and of the utility for the attributes of the choice alternatives. It is plausible that the design, preferences and ANA might be interrelated in more complex ways. Should this be the case, then different conclusions could be drawn. Certainly, choices could be simulated based on a more complex formulation of ANA, for example one which is more behaviourally motivated, and the results similarly interpreted. This remains an area for future work, likely in tandem with the development of these alternative ANA frameworks. The findings of this chapter nonetheless demonstrate that the degree of bias induced by ANA may not be trivial.

In conclusion, this chapter has demonstrated the biasing influence of ANA on the mean and dispersion of random parameters, and explored the interactions between ANA and sign violation. Two techniques that have been proposed for handling ANA, the censored normal distribution and the IANA model, were tested for their ability to recover the true parameters and ANA rates[12]. Whilst they performed well in very specific cases, it was found that they did not do so when tested across a range of scenarios, thus limiting the applicability of these methods. These shortcomings provide an impetus for the development of a technique that is more robust, namely, the RPANA model. The flexibility of the RPANA model means that, theoretically, it should be more robust, since unlike the IANA and CANA models it can accommodate ANA *and* preference heterogeneity, and unlike with the censored normal distribution, these two behaviours are captured with separate parameters.

[12] In the case of the IANA model, the mean of the RP distribution.

Chapter 5

Evaluating the performance of the random parameters attribute nonattendance model with simulations

5.1 Introduction

This chapter tests the ability of the RPANA model to accurately recover ANA rates, in a controlled environment, on simulated data. The estimated RP distributions, which represent preference heterogeneity amongst those that attend to the attributes, are also compared with the true distributions, to see if any bias is introduced.

The use of simulated data allows the performance to be tested under ideal circumstances, where the true distribution conveniently matches a distribution that can readily be specified by the analyst. Real empirical applications are likely to have a more complex pattern of preference heterogeneity, which the estimated distribution can only approximate. Nonetheless, it has been demonstrated in Chapter 4 that existing approaches, such as the IANA model, fail to satisfactorily represent even very simple distributions[1]. Indeed, this is one of the key motivations for the development of the RPANA model. Adequately capturing these simple distributions would be a good starting point for the RPANA model.

Only RPIANA models will be tested[2]. That is, the models will assume that ANA is independent across the attributes. This aligns with the manner in which the simulated data

[1] Whilst the IANA and CANA models cannot represent a distribution of preferences, the estimates of the means of these distributions, and the ANA rates, are biased with these models.

[2] But will be referred to as RPANA models herein for brevity.

were generated. Correlations in nonattendance across attributes could be introduced in the simulated data, to test the performance of the both the RPCANA and RPIANA models in the presence of such correlation, however, this will remain an area for future research.

5.2 Methodology

The methodology for this chapter largely draws upon that employed in Chapter 4. The same datasets were used, where the data generation process was documented in Section 4.2.1. However, the first three treatments were not used, for they were generated with full attendance across all attributes, and there is little point estimating ANA with the RPANA model when there is no ANA[3]. Additionally, RPANA models were estimated on only 30 of the 100 datasets. This was purely for practical reasons, as with additional layers of integration, the RPANA model is slower to estimate than the RPL model.

Again, 5000 Halton draws were used for the random parameters. As in Chapter 4, we are interested in parameter recovery when the same distribution is applied that was used to generate the dataset, even though the true distribution is latent in empirical contexts. Datasets were generated with the normal, triangular, uniform, lognormal, and censored normal distributions. However, it was found that the only distributions with which the RPANA model could be estimated are those that are constrained to be of the same sign, across their entire domain. This prevented estimation of the normal, triangular, and uniform distributions. Additionally, identification problems were observed with the censored normal distribution. A full discussion of these problems, which were evident with both simulated and empirical data, was provided in Section 3.3.2. This left only the lognormal distribution, which indeed is a distribution that is widely used for its ability to impose a constraint on the sign of the coefficients.

However, the triangular distribution can be constrained to prevent sign violations, by fixing the spread to be a multiple, between zero and one, of the mean. A triangular distribution with the spread constrained to equal the mean (i.e., a multiple of one) was used with the RPANA model, on the datasets generated with the triangular distribution. One third of the treatments were generated with the spread equal to the mean, and so those distributions aligned with the constraint that was imposed during estimation. Ideally, those distributions would be estimated without bias. However, for the remaining treatments in

[3]The exception being a test to see if ANA is estimated by the RPANA model when it does not exist. Limited testing on these treatments revealed very low rates of ANA for just some of the datasets.

which the spread does not equal the mean, the constraint imposed is likely to lead to some bias. Other constraints, such as a multiple between zero and one, could be imposed, but were not. This will be discussed later in this chapter. A constrained uniform distribution was also tested on the datasets generated with the uniform distribution, but estimation was not stable, suggesting a possible identification problem (again refer to Section 3.3.2 for a discussion on this point). Chapter 6 employs the Rayleigh distribution, which imposes a sign constraint, however, this distribution will not be tested with simulated data[4]. Thus, the two final distributions employed are the lognormal and the constrained triangular.

For each of the attributes, some of the treatments specify complete attendance to that attribute. Modelling nonattendance to these attributes in most cases resulted in a zero percent ANA rate, which signifies an accurate recovery of the true ANA rate[5]. This is represented by the parameter controlling ANA, γ_{c_a}, approaching negative infinity. This frequently led to instability in model estimation, and so ANA was not modelled for those attributes in those treatments.

5.3 Random parameters attribute nonattendance model results

5.3.1 Constrained triangular distribution

Where the spread equals the mean in the true distribution, the RPANA model performs well, as evidenced in Table 5.1. The true ANA rates are not estimated with complete accuracy, but the magnitude of the differences are small, and the estimated rate is consistently a slight underestimate. When the spread is less than the magnitude of the mean in the true distribution, the estimated ANA rates are more severely underestimated. The downward bias appears to be fairly absolute in magnitude, with little absolute variation as the true ANA rate varies, and consequently a very large variation in relative terms. Consider, for example, the average ANA rates for the first attribute, in treatments with the narrowest spread ($\sigma = 0.1$) as the true ANA rate varies. In treatment eight, with a true ANA rate of 15 percent, the estimated ANA rate is six percent, which represents an underestimate of nine percent. In treatment ten, with a true ANA rate of 30 percent, the underestimate is 9.7 percent; and in treatment 18, with a true ANA rate of 45 percent, the underestimate is 10 percent.

[4]Extending the simulations to include the Rayleigh distribution will remain an area for future research.

[5]There were, however, a small number of cases of 'false positives', that is, estimations of nonzero ANA rates when the true rate was zero.

Table 5.1: RPANA models estimated on datasets with triangularly distributed true sensitivities

Tr.	Parameter One				Parameter Two				Parameter Three			
	Actual[1]		Estimated		Actual[2]		Estimated		Actual[3]		Estimated	
	σ	NA (%)	$\bar{\mu}/\bar{\sigma}$	NA (%)	σ	NA (%)	$\bar{\mu}/\bar{\sigma}$	NA (%)	σ	NA (%)	$\bar{\mu}/\bar{\sigma}$	NA (%)
4	0.2	30	-0.285	23.1	0.06	45	-0.058	42.6	0.9	30	0.886	29.0
5	0.3	30	-0.305	27.7	0.02	45	-0.051	32.7	0.3	30	0.791	19.5
6	0.1	30	-0.272	19.8	0.04	45	-0.054	36.6	0.6	30	0.823	22.8
7	0.3	15	-0.303	13.2	0.06	15	-0.059	13.8	0.3	45	0.769	35.1
8	0.1	15	-0.276	6.0	0.02	15	-0.053	3.5	0.6	45	0.799	38.0
9	0.2	15	-0.286	8.7	0.04	15	-0.055	7.4	0.9	45	0.871	44.0
10	0.1	30	-0.273	20.3	0.04	0	-0.060	-	0.6	45	0.811	38.7
11	0.2	30	-0.286	23.4	0.06	0	-0.060	-	0.9	45	0.880	44.8
12	0.3	30	-0.306	28.1	0.02	0	-0.059	-	0.3	45	0.783	36.2
13	0.3	15	-0.305	13.5	0.02	30	-0.053	18.3	0.9	15	0.884	13.2
14	0.1	15	-0.276	5.5	0.04	30	-0.055	22.2	0.3	15	0.795	3.6
15	0.2	15	-0.286	8.2	0.06	30	-0.059	29.0	0.6	15	0.822	6.6
16	0.2	45	-0.281	38.1	0.06	45	-0.059	43.4	0.6	0	0.889	-
17	0.3	45	-0.303	42.8	0.02	45	-0.052	34.0	0.9	0	0.900	-
18	0.1	45	-0.267	35.0	0.04	45	-0.054	36.9	0.3	0	0.881	-
19	0.3	45	-0.305	43.5	0.04	0	-0.060	-	0.6	0	0.883	-
20	0.1	45	-0.266	35.1	0.06	0	-0.060	-	0.9	0	0.901	-
21	0.2	45	-0.283	38.7	0.02	0	-0.060	-	0.3	0	0.878	-
22	0.3	45	-0.305	43.7	0.04	30	-0.055	22.2	0.9	45	0.886	44.9
23	0.1	45	-0.267	35.6	0.06	30	-0.059	28.4	0.3	45	0.778	35.4
24	0.2	45	-0.282	38.6	0.02	30	-0.053	17.3	0.6	45	0.816	38.8
25	0.1	0	-0.293	-	0.06	45	-0.059	43.6	0.3	15	0.770	3.1
26	0.2	0	-0.300	-	0.02	45	-0.051	32.9	0.6	15	0.812	6.6
27	0.3	0	-0.310	-	0.04	45	-0.054	36.4	0.9	15	0.886	13.2
28	0.2	0	-0.299	-	0.04	30	-0.055	22.2	0.3	30	0.766	18.4
29	0.3	0	-0.308	-	0.06	30	-0.059	28.1	0.6	30	0.812	22.0
30	0.1	0	-0.294	-	0.02	30	-0.052	17.3	0.9	30	0.868	28.5
31	0.2	30	-0.282	22.5	0.02	15	-0.053	3.2	0.3	15	0.800	4.5
32	0.3	30	-0.305	27.8	0.04	15	-0.055	7.3	0.6	15	0.828	7.7
33	0.1	30	-0.272	19.9	0.06	15	-0.059	13.4	0.9	15	0.884	13.1
34	0.1	15	-0.276	5.8	0.02	15	-0.054	3.7	0.9	30	0.880	28.4
35	0.2	15	-0.285	8.1	0.04	15	-0.055	6.4	0.3	30	0.784	19.1
36	0.3	15	-0.303	12.9	0.06	15	-0.059	13.5	0.6	30	0.822	23.2

1. μ =-0.3 for all treatments. 2. μ =-0.06 for all treatments. 3. μ=0.9 for all treatments.

When the spread equals the mean in the true distribution, both these values are estimated[6] with a high degree of accuracy, with only a slight downward bias for the third attribute. As the true spread decreases, a mild downward bias on the mean becomes evident. The bias in the spread is much more prominent. So long as the spread is constrained to equal the mean for estimation, the two measures will be confounded when the two values are not equal in the true distribution. Unsurprisingly, the mean is the dominant moment, and so most bias is instead introduced into the spread. Again, the downward bias, this time in the mean, appears not to vary in absolute terms, as the true ANA rate varies.

The estimated ANA rates and means documented in Table 5.1 are averaged across the 30 models runs per treatment. In some runs, the estimated ANA rate was zero, when the true rate was nonzero. This only occurred for the second and third attribute, in treatments in which the true ANA rate was 15 percent, when the true spread was either one third or, less frequently, two thirds of the mean. In such treatments, the constraint imposed on the triangular distribution is forcing there to be mass near zero that is not reflective of the true distribution. Since the true ANA rate is only 15 percent, it is likely that for these runs, the nonattenders are being represented by the mass near zero enforced by the constraint, allowing the ANA point mass to collapse. The enforced mass appears to not be sufficiently large to capture ANA rates of either 30 or 45 percent in their entirety, although it appears to capture *some* of the ANA.

Recall from Section 4.4.2 in the previous chapter that when the IANA model was estimated on simulated data generated with triangularly distributed preference heterogeneity, the accuracy of the estimated ANA rate increased as the true spread reduced to zero. This is plausible, as with random parameter spreads of zero, the RPL model will collapse to an MNL model, which is nested within the IANA model. Conversely, as the constraint on the triangular distribution has been specified, the accuracy of the ANA rates and sensitivities increases as the true spread approaches the magnitude of the mean from below. Where there is bias, the IANA model is overestimating the ANA rates and the magnitudes of the means, while the RPANA model is underestimating the ANA rates and the magnitudes of the means.

Much of the bias in the RPANA model appears to be stemming from the constraint imposed upon the triangular distribution, which is necessary to prevent sign violations and so lessen the chance of identification problems. However, if the spread is constrained to be some other multiple of the mean, between zero and one, sign violations will also be prevented, and distributions may be accurately estimated that were generated with the spread defined

[6]With a single parameter due to the constraint imposed on the triangular distribution.

as some multiple of the mean other than one. Models specified with other multiples were estimated, without identification problems, but are not reported here. It should be noted that if this approach is employed, then the analyst needs to make a decision not just on the distribution employed, but also the constraint imposed. When estimating such a model with empirical data, the true ANA and taste parameter values are known, so various constraints on the triangular distribution could be tested, and the specification leading to the best model fit accepted. As already noted, the true distribution is unlikely to closely conform to *any* common distribution, however, some form of triangular distribution may provide an adequate approximation.

5.3.2 Lognormal distribution

Considerable problems were encountered when attempting to model nonattendance to the third attribute with the RPANA model, for the datasets generated with the lognormal distribution. A large percentage of the models failed to converge, with ANA rates frequently tending to zero. It is believed that this was a consequence of the considerable right skew exhibited by the lognormal distribution specified for third attribute. This was illustrated in Figure 4.1 (p.98) in the previous chapter. The right skew places a considerable mass near zero, so that a moderate percentage of the coefficients under the continuous distribution represent a very low sensitivity to the attribute. This may lead to confounding between ANA and low sensitivity. The continuous distribution may readily 'absorb' the ANA mass generated in the simulations, especially when the ANA rates are low. The results from the IANA model estimated on these datasets, which were detailed in Section 4.4.4, provide further evidence. The nonattendance rates for the third attribute were extremely high, up to 80.9 percent in those treatments with a true ANA rate of 45 percent. This suggests that the estimated point mass at zero is representing a large proportion of the true continuous distribution, where this is likely due to the proximity of these coefficients to zero. Now that a continuous distribution is estimated with the RPANA model, in place of a point estimate, this continuous distribution can capture the ANA, to the extent that the ANA point mass collapses. Consequently, the RPANA models will be specified without modelling nonattendance to the third attribute. Instead, it is simply acknowledged that in most instances, the ANA cannot be separately identified for this attribute. This is a cautionary note, that a large mass of low sensitivities to an attribute may preclude the estimation of ANA to that attribute.

The results for the RPANA models with the lognormal distribution are presented in Table

5.2 for the first attribute, and Table 5.3 for the second. The ANA rate is recovered fairly accurately for the first attribute, with an average error of 3.3 percent, and no discernable bias, with the recovered ANA rates variously slightly above and slightly below the true rates. By contrast, the estimated rates of nonattendance to the second attribute are consistently lower than the true rates, by an average of 10.4 percent. The IANA results from Section 4.4.4 were different again, where that model consistently and extensively overestimated the ANA rate. Therefore, while the accuracy of the ANA rates is not consistent over the three distributions tested with the RPANA model (no bias and small error, mild downward bias, and a collapse to zero percent ANA with extensive estimation problems, for the first, second and third attributes respectively), it is more promising than the wildly inaccurate rates recovered with the IANA model. Further, a relatively high level of estimation accuracy is retained over a much larger range of true values than with the IANA model. Zero percent ANA rates were estimated in a small percentage of model runs, where the true ANA rate was not zero. Unlike with the triangular distribution, there was no discernable pattern as to which treatments this occurred in.

The biases of the various measures of the lognormal distribution (as opposed to the underlying normal distribution) are somewhat inconsistent, and the focus will be on the mean. For the first attribute, the mean is consistently underestimated, by an average of 4.6 percent. In contrast, the bias varies across treatments for the second attribute, but there is an average overestimation of 3.7 percent. Nonetheless, these biases are mild, and contrast with the stronger and more consistent overestimation of the mean observed with the IANA model in Section 4.4.4.

5.4 Discussion

Overall, the RPANA model is found to perform strongly on the simulated datasets, in terms of its ability to recover true distributions that contain both a continuous component, and a point mass at zero. In part, the accuracy depends on the ability of the estimated distribution to match the real distribution. The limits imposed on the RP distributions, to facilitate identification of the model, may limit this alignment to some extent. However, some distributions may act as reasonable substitutes. The peaked and symmetrical nature of the triangular distribution, for example, may be an adequate approximation of the normal distribution, although it will not capture outliers as effectively.

The simulations have shown that the ANA point mass and the continuous distribution may

Table 5.2: RPANA models estimated on datasets with lognormally distributed true sensitivities, parameter one

Tr.	Actual[1]				Estimated					Proportion				
	Med.	**Mode**	**S.D.**	**NA (%)**	**Mean**	**Med.**	**Mode**	**S.D.**	**NA (%)**	**Mean**	**Med.**	**Mode**	**S.D.**	**NA**
4	-0.262	-0.200	0.208	30	-0.281	-0.251	-0.201	0.191	28.7	-0.063	-0.041	0.006	-0.085	-0.043
5	-0.238	-0.150	0.229	30	-0.272	-0.226	-0.156	0.199	29.0	-0.092	-0.051	0.037	-0.131	-0.035
6	-0.282	-0.250	0.193	30	-0.288	-0.276	-0.253	0.183	29.4	-0.039	-0.023	0.010	-0.055	-0.020
7	-0.238	-0.150	0.229	15	-0.286	-0.241	-0.171	0.206	16.8	-0.047	0.012	0.140	-0.102	0.122
8	-0.282	-0.250	0.193	15	-0.294	-0.281	-0.255	0.187	15.7	-0.018	-0.006	0.019	-0.030	0.050
9	-0.262	-0.200	0.208	15	-0.290	-0.261	-0.212	0.196	16.0	-0.032	-0.003	0.059	-0.061	0.067
10	-0.282	-0.250	0.193	30	-0.291	-0.278	-0.254	0.185	29.8	-0.029	-0.015	0.014	-0.043	-0.005
11	-0.262	-0.200	0.208	30	-0.283	-0.254	-0.205	0.192	29.6	-0.055	-0.030	0.024	-0.080	-0.014
12	-0.238	-0.150	0.229	30	-0.276	-0.231	-0.161	0.200	30.1	-0.079	-0.030	0.076	-0.126	0.002
13	-0.238	-0.150	0.229	15	-0.286	-0.239	-0.167	0.207	16.6	-0.048	0.003	0.113	-0.097	0.107
14	-0.282	-0.250	0.193	15	-0.293	-0.280	-0.255	0.187	15.2	-0.022	-0.008	0.019	-0.035	0.015
15	-0.262	-0.200	0.208	15	-0.288	-0.259	-0.209	0.195	15.2	-0.039	-0.012	0.043	-0.064	0.014
16	-0.262	-0.200	0.208	45	-0.284	-0.256	-0.207	0.192	43.7	-0.052	-0.024	0.036	-0.080	-0.029
17	-0.238	-0.150	0.229	45	-0.282	-0.240	-0.174	0.202	45.7	-0.059	0.008	0.157	-0.121	0.016
18	-0.282	-0.250	0.193	45	-0.286	-0.271	-0.244	0.183	44.3	-0.046	-0.038	-0.022	-0.054	-0.017
19	-0.238	-0.150	0.229	45	-0.282	-0.238	-0.170	0.203	45.3	-0.059	0.000	0.130	-0.115	0.008
20	-0.282	-0.250	0.193	45	-0.290	-0.276	-0.250	0.185	44.3	-0.032	-0.021	0.000	-0.042	-0.016
21	-0.262	-0.200	0.208	45	-0.288	-0.261	-0.214	0.193	45.2	-0.039	-0.005	0.068	-0.072	0.004
22	-0.238	-0.150	0.229	45	-0.277	-0.234	-0.168	0.198	44.2	-0.077	-0.016	0.119	-0.135	-0.017
23	-0.282	-0.250	0.193	45	-0.286	-0.273	-0.249	0.182	43.3	-0.045	-0.032	-0.005	-0.058	-0.039
24	-0.262	-0.200	0.208	45	-0.284	-0.257	-0.211	0.190	44.0	-0.054	-0.019	0.055	-0.088	-0.023
25	-0.282	-0.250	0.193	0	-0.294	-0.278	-0.249	0.188	-	-0.021	-0.016	-0.006	-0.025	-
26	-0.262	-0.200	0.208	0	-0.289	-0.256	-0.200	0.198	-	-0.036	-0.024	0.002	-0.049	-
27	-0.238	-0.150	0.229	0	-0.286	-0.230	-0.150	0.215	-	-0.047	-0.032	-0.002	-0.062	-
28	-0.262	-0.200	0.208	0	-0.289	-0.255	-0.198	0.199	-	-0.037	-0.027	-0.008	-0.046	-
29	-0.238	-0.150	0.229	0	-0.287	-0.232	-0.152	0.215	-	-0.045	-0.027	0.011	-0.063	-
30	-0.282	-0.250	0.193	0	-0.293	-0.277	-0.248	0.188	-	-0.025	-0.020	-0.009	-0.030	-
31	-0.262	-0.200	0.208	30	-0.283	-0.254	-0.205	0.191	29.8	-0.056	-0.029	0.027	-0.082	-0.005
32	-0.238	-0.150	0.229	30	-0.275	-0.227	-0.154	0.202	29.2	-0.083	-0.048	0.027	-0.117	-0.028
33	-0.282	-0.250	0.193	30	-0.288	-0.274	-0.248	0.183	29.1	-0.041	-0.030	-0.006	-0.053	-0.031
34	-0.282	-0.250	0.193	15	-0.293	-0.280	-0.255	0.186	15.4	-0.022	-0.009	0.020	-0.036	0.028
35	-0.262	-0.200	0.208	15	-0.289	-0.261	-0.213	0.194	15.9	-0.037	-0.004	0.065	-0.069	0.062
36	-0.238	-0.150	0.229	15	-0.284	-0.238	-0.167	0.206	16.0	-0.053	-0.001	0.111	-0.102	0.070

1. *mean* =-0.3 for all treatments.

Table 5.3: RPANA models estimated on datasets with lognormally distributed true sensitivities, parameter two

Tr.	Actual[1]				Estimated					Proportion				
	Med.	Mode	S.D.	NA (%)	Mean	Med.	Mode	S.D.	NA (%)	Mean	Med.	Mode	S.D.	NA
4	-0.050	-0.035	0.044	45	-0.063	-0.051	-0.034	0.046	41.3	0.043	0.023	-0.018	0.065	-0.083
5	-0.058	-0.055	0.037	45	-0.055	-0.051	-0.045	0.036	36.9	-0.088	-0.542	-0.724	-0.240	-0.180
6	-0.055	-0.045	0.040	45	-0.058	-0.051	-0.039	0.040	38.4	-0.030	-0.546	-0.746	-0.186	-0.146
7	-0.050	-0.035	0.044	15	-0.064	-0.055	-0.041	0.045	15.8	0.071	-0.163	-0.282	-0.023	0.053
8	-0.058	-0.055	0.037	15	-0.060	-0.059	-0.056	0.038	13.6	0.007	-0.185	-0.282	-0.076	-0.094
9	-0.055	-0.045	0.040	15	-0.062	-0.058	-0.049	0.041	14.1	0.042	-0.172	-0.282	-0.046	-0.058
10	-0.055	-0.045	0.040	0	-0.062	-0.057	-0.049	0.041	-	0.034	0.030	0.059	0.001	-
11	-0.050	-0.035	0.044	0	-0.063	-0.053	-0.039	0.045	-	0.047	0.047	0.082	0.014	-
12	-0.058	-0.055	0.037	0	-0.062	-0.060	-0.056	0.039	-	0.030	-0.007	-0.029	0.016	-
13	-0.058	-0.055	0.037	30	-0.060	-0.058	-0.054	0.038	27.7	0.004	-0.371	-0.533	-0.153	-0.078
14	-0.055	-0.045	0.040	30	-0.060	-0.055	-0.046	0.040	27.1	0.008	-0.370	-0.546	-0.126	-0.097
15	-0.050	-0.035	0.044	30	-0.062	-0.052	-0.037	0.045	27.8	0.041	-0.375	-0.583	-0.064	-0.074
16	-0.050	-0.035	0.044	45	-0.061	-0.049	-0.031	0.046	37.8	0.016	-0.553	-0.782	-0.084	-0.160
17	-0.058	-0.055	0.037	45	-0.052	-0.048	-0.041	0.034	33.2	-0.128	-0.542	-0.727	-0.231	-0.263
18	-0.055	-0.045	0.040	45	-0.055	-0.048	-0.036	0.038	41.4	-0.087	-0.540	-0.747	-0.162	-0.079
19	-0.055	-0.045	0.040	0	-0.062	-0.057	-0.047	0.041	-	0.034	0.039	0.063	0.016	-
20	-0.050	-0.035	0.044	0	-0.063	-0.053	-0.038	0.045	-	0.049	0.051	0.087	0.015	-
21	-0.058	-0.055	0.037	0	-0.061	-0.060	-0.056	0.038	-	0.025	-0.006	-0.032	0.021	-
22	-0.055	-0.045	0.040	30	-0.060	-0.054	-0.044	0.041	26.2	0.004	-0.364	-0.551	-0.099	-0.128
23	-0.050	-0.035	0.044	30	-0.062	-0.051	-0.035	0.045	26.0	0.026	-0.364	-0.566	-0.068	-0.132
24	-0.058	-0.055	0.037	30	-0.057	-0.054	-0.048	0.037	23.4	-0.044	-0.367	-0.532	-0.145	-0.218
25	-0.050	-0.035	0.044	45	-0.058	-0.046	-0.030	0.044	36.8	-0.035	-0.565	-0.791	-0.094	-0.182
26	-0.058	-0.055	0.037	45	-0.058	-0.055	-0.051	0.037	40.1	-0.032	-0.545	-0.724	-0.249	-0.108
27	-0.055	-0.045	0.040	45	-0.060	-0.054	-0.043	0.040	41.1	-0.002	-0.555	-0.760	-0.174	-0.087
28	-0.055	-0.045	0.040	30	-0.061	-0.056	-0.048	0.040	27.7	0.016	-0.369	-0.547	-0.122	-0.075
29	-0.050	-0.035	0.044	30	-0.062	-0.052	-0.037	0.045	26.9	0.032	-0.376	-0.585	-0.062	-0.105
30	-0.058	-0.055	0.037	30	-0.058	-0.056	-0.052	0.037	26.9	-0.026	-0.373	-0.531	-0.161	-0.104
31	-0.058	-0.055	0.037	15	-0.061	-0.059	-0.056	0.038	13.5	0.015	-0.180	-0.277	-0.070	-0.103
32	-0.055	-0.045	0.040	15	-0.062	-0.057	-0.048	0.041	14.5	0.041	-0.171	-0.281	-0.043	-0.030
33	-0.050	-0.035	0.044	15	-0.064	-0.054	-0.039	0.045	14.9	0.063	-0.166	-0.286	-0.024	-0.003
34	-0.058	-0.055	0.037	15	-0.060	-0.058	-0.055	0.037	12.3	-0.005	-0.187	-0.285	-0.075	-0.177
35	-0.055	-0.045	0.040	15	-0.062	-0.057	-0.049	0.040	14.0	0.029	-0.172	-0.275	-0.054	-0.065
36	-0.050	-0.035	0.044	15	-0.063	-0.054	-0.040	0.045	14.7	0.057	-0.165	-0.287	-0.022	-0.020

1. *mean* =-0.06 for all treatments.

become confounded, leading to erroneous model outputs. This may occur due to properties of the true distribution, such as a large mass of coefficients with low magnitude, as with the third attribute in the lognormal simulations. Alternatively, it may be the consequence of a decision made by the analyst, such as the enforcement of a constraint on a distribution. This was evident when the spread was constrained to equal the mean with the triangular distribution. As mass was forced to exist near zero, this approximated ANA, and led to the collapse of the ANA point mass, even though one existed in the true distribution.

A lesson from this second point in particular is that a zero percent ANA rate should not be taken at face value, if a range of distributions have not been tested. For example, Hess et al. (2011) implemented a form of RPIANA model, but only presented models specified with the lognormal distribution. It is unknown whether they tested other distributions, but it is recommended that this be done. Of course, decisions regarding model specification are much harder when the true distribution is unknown. In particular, the true taste distribution may not have a close counterpart in the range of distributions available to the analyst. Nonetheless, the testing of as many distributions as possible is suggested, especially since a consequence of not doing so may be erroneous inferences about ANA rates.

This chapter has demonstrated that the RPANA model shows promise as a means of capturing ANA and preference heterogeneity, even if some caution is warranted. The RPANA model performs better over a wider range of ANA and preference heterogeneity configurations than the IANA model and the RPL model with censored normal distributions. The next chapter will evaluate the performance of the RPANA model in an empirical setting.

Chapter 6

Application of the random parameters attribute nonattendance model to an air travel stated choice experiment

6.1 Introduction

This chapter tests the RPANA model on an empirical dataset. Specifically, the choice of a short haul airline ticket between Sydney and Melbourne is examined. A large number of ANA and RPL models are first explored, and serve as a reference point. Then, various specifications of the RPANA model are compared, both to each other, and to the reference models, to gain a nuanced understanding of the strengths and weaknesses of the RPANA model. Additionally, the chapter provides insights into ANA behaviour in the context of air travel choice.

First, applications of discrete choice models to air travel behaviour are reviewed in Section 6.2. Next, Section 6.3 introduces the empirical setting for this chapter. Section 6.4 provides the MNL model results. Whilst Chapter 4 demonstrated with simulations that taste heterogeneity is likely to bias the ANA model, this model will also be tested empirically, in Section 6.5. The ANA model serves as an important reference point for the RPANA model, as, subject to the same assumptions regarding independence of ANA, the ANA model is nested within the RPANA model. The ANA rates from the ANA model will be considered for their reasonableness, and be one point of comparison with the RPANA model. Models that variously handle ANA to one and all attributes will be compared, as will the IANA and CANA specifications of the model, where no previous studies have compared these models.

Recall that the IANA model assumes ANA to be independent across attributes, while the CANA model allows for correlation across attributes. The conditional parameter estimates of the ANA model will be examined, to provide some further insight into what exactly the ANA model is capturing.

The RPL model is also nested within the RPANA model (recall Section 3.3.1), and serves as another reference point. Section 6.6 explores RPL models with many different RP distributions, where some of the same distributions are also later specified for the RPANA model, so as to investigate the impact of the ANA point mass on the distribution. Insights are also drawn into how the various distributions in the RPL model may be approximating ANA. Also, the censored normal distribution is an alternative mechanism for capturing ANA, and so RPL models with censored normal distributions will serve as a reference point in which ANA can be handled.

To gauge the performance of the RPANA model, alternative specifications are systematically investigated. RPANA models with ANA modelled for just one attribute, using a variety of distributions, allow a clear comparison to be made with their RPL counterparts (Section 6.7.2). In Section 6.7.3, the stated ANA responses are used as covariates in the the ANA assignment models. Next, based on building evidence that ANA is not independent for fare and flight time, a RPANA model is estimated that confirms this, and which shows that for these attributes the independence assumption is detrimental to the performance of the model. To investigate the independence assumption further, Section 6.7.5 compares RPIANA and RPCANA models in which ANA is handled for two attributes only. Section 6.7.6 presents a hybrid RPANA model that handles ANA for all attributes, and assumes independence of ANA across just some groups of attributes. Section 6.7.7 compares this model with a RP-CANA model in which correlation in ANA can be handled between all attributes. A brief discussion is provided in Section 6.8, although a more extensive critique of the RPANA model is reserved for the next chapter.

6.2 Discrete choice models of air travel behaviour

Discrete choice modelling methodologies have been used to investigate a wide range of air travel behaviour. Many studies have relied on revealed preference data. For example, Kanafani and Sadoulet (1977) modelled the choice among fare types for long haul journeys, Theis et al. (2006) investigated the impact of connection time at hubs, and Coldren et al. (2003) modelled aggregate air travel itinerary shares for all city-pairs in the United

States. Proussaloglou and Koppelman (1995) examined the choice of airline for recent trips, considering influences such as market share, frequent flyer membership, and perception of fare and service level.

Stated choice techniques have also been used extensively as a tool for investigating air travel behaviour. They have proven to be particularly useful in situations where certain attribute levels or combinations of attribute levels do not exist in the market. Hensher and Louviere (1983) investigated the influence of ticket attributes on choice. Indeed, it was one of the earliest stated preference studies concerning products decomposed into multiple attributes, although they relied on rankings rather than choice data. Stated choice studies have considered a range of specific choices related to air travel, including choice of airline ticket only (Rose et al., 2005; Collins et al., 2012), choice of airline and ticket class (Hensher et al., 2001), combined airport and airline choice (Bradley, 1998; Hess et al., 2007a; Hess, 2007), and choice between air travel and other modes (Ortúzar and Simonetti, 2008).

A number of modelling, survey and other empirical issues have been considered. Warburg et al. (2006) modelled systematic and random preference heterogeneity in a study of itinerary choice, and stressed that systematic sources of preference heterogeneity should be handled first. Bliemer and Rose (2011) examined the impact of different types of experimental designs on airline choice in SC studies. Attribute nonattendance has previously been applied in the air travel behaviour literature, in the work of Rose et al. (2005). Various alternative elicitation methods have been employed, as variants on the traditional SC presentation and response mechanisms. Proussaloglou and Koppelman (1999) completed a study whereby respondents were presented with a travel scenario, and were required to elicit from the interviewer the available flights as described by schedule and fare. Flights could be revealed in any order the respondent wished, according to schedule or fare, and a choice could be made at any stage. Collins et al. (2012) presented respondents with a choice mechanism that mimicked that of an online travel agent. The choice tasks contained a realistically large number of flight alternatives, each described by a large number of attributes. Respondents could filter the amount of information using search tools that eliminated alternatives, and tools for hiding and showing attributes. Indeed, considering the large number of alternatives and attributes in many air travel choices, ANA is as a very plausible behavioural response, and handling this ANA might be particularly important in these choice contexts.

The empirical study in this chapter considers the choice of air ticket only, in a conventional SC choice task. In the study, there is little scope for systematically handling preference heterogeneity, due to a very limited number of potential covariates, with little variability across those that exist. Hence, the study will rely extensively on random preference heterogeneity.

6.3 Empirical setting

The empirical setting for this chapter is an SC experiment conducted in early 2004, that was based on a short haul flight between Sydney and Melbourne, Australia. Respondents were asked to imagine that they were making the flight for holiday travel. Each choice task contained three labelled flight alternatives. A choice was made between one flight each from three airlines: Qantas, the dominant Australian carrier; Virgin Blue, then a relatively young airline with four years of operations; and Air New Zealand (Air NZ), a foreign carrier that does not operate the Sydney-Melbourne route. A fourth, no-choice option was presented, which signalled that the respondent would not want to make any of the three flights. Two choices were obtained: one that included the no-choice alternative, and a forced choice over the three airlines. The analysis contained herein makes use only of the forced choice.

Each alternative was described by four attributes: fare, flight time, departure time, and flight time variability. Fare assumed one of four levels in Australian dollars: \$79, \$99, \$119 and \$139. Flight time was either 40, 50, 60 or 70 minutes. Departure time was either 6am, 10am, 2pm or 6pm. Flight time variability was used to convey the range of likely flight times, with the flight time attribute level the expected flight time, and the flight time variability attribute providing a symmetric disturbance around the mean. The flight time variability level was calculated as a percentage of that alternative's flight time, with levels of ±5, ±7.5, ±10, and ±12.5 percent. For example, if the flight time was 60 minutes, and the variability level was 10 percent, the flight time variability would be ±6 minutes, and the respondent could expect the flight to be between 54 and 66 minutes in duration. The introduction of a flight time variability attribute was motivated to a large extent by the frequent delays on the busy Sydney-Melbourne route. However, the attribute was not well received, with 69 percent of respondents stating in a subsequent question that they ignored the attribute. Tests with random parameters imply an even distribution of respondents for and against flight time variability, suggesting that many did not understand the attribute. It will be omitted from subsequent analysis as the ambiguity just adds unnecessary heterogeneity.

Each airline alternative was described by the same set of attribute levels that were varied via an orthogonal experimental design. That is, no airline had a disproportionate number of each of the attribute levels, despite, for example, a tendency for Virgin Blue to offer cheaper tickets than Qantas in the market. The orthogonal design contained 40 choice tasks in total, all of which were completed by 213 respondents, in one of three ways. As a part of a broader research agenda investigating multiple survey sessions per respondent spread over time, respondents either completed all 40 choice tasks in one sitting; 20 choice tasks each in two sessions, with one week of separation; or 10 choice tasks per session over four sessions each separated by a week. Regardless of the configuration employed, this chapter utilises the first 20 choice tasks completed by each respondent. The sample consisted of students, with an average age of 21, minimum of 18, maximum of 38, and standard deviation of 2.2 years. Fifty nine percent of the sample was female, and 41.43 percent had made a holiday trip to Melbourne prior to the study. No other socio-demographic or experience information was gathered.

An examination of the 213 respondents revealed two who chose the Qantas alternative for all 20 choice tasks, and one who always chose the Virgin Blue alternative. Since one flight from each airline was presented in each choice task, if this is a true representation of lexicographic choice behaviour, then no trading is occurring between the attributes describing the airlines. It may be that trading is taking place across the airline label and the attributes, with the attributes in the other two alternatives just failing to compensate in each of 20 successive choices. Nonetheless, the length of the panel suggests this is unlikely[1], and so these three observations are dropped. Interestingly, this is an extreme case of ANA, where all attributes are ignored, and only the airline labels are attended to.

Additionally, 55 respondents were identified as always choosing an alternative with the lowest fare. This is not as problematic as with lexicographic choice based on airline, as in many choice tasks, two or more alternatives shared the lowest fare, and in these cases the other attributes can be used by the respondent to determine the favoured alternative. Specifically, 16 choice tasks in the experimental design of 40 choice tasks had a tie in fare across two alternatives, and one alternative had a tie across all three alternatives. As further evidence that these 55 respondents who always chose the lowest fare might have also been considering other attributes, consider their responses to questions on whether they ignored each of the attributes. Thirty three stated that they considered both flight time and departure time, 13 ignored departure time only, six ignored flight time only, and only one stated that they ignored both departure time and flight time. Whilst these responses cannot be relied upon

[1]Panel lengths in SC experiments are commonly shorter. Bliemer and Rose (2011) examined tier one transportation journals from January 2000 to August 2009, and found an average panel length of 9.4 and a median length of nine. The shorter the panel, the less confidence can be placed on any interpretation of lexicographic behaviour.

as completely truthful, they do suggest that it is likely that a majority of the 55 respondents are considering more than fare alone. Consequently, these 55 respondents are retained in the sample. The final sample size is 4200 observations across 210 respondents.

Of these 210 respondents, 6.9 percent stated that they ignored fare, 18.1 percent flight time, and 15.95 percent departure time. Respondents may also have ignored the airline label, however, they were not asked if this was the case. Whilst the literature has called into question the reliability of the responses to these questions (Hess and Rose, 2007), and cautions against using them deterministically (Hensher et al., 2007), they nonetheless provide a broad sense of what the nonattendance rates might be in aggregate across respondents. Further, by suggesting that there is likely to be at least some incidence of ANA in this dataset, they motivate the analyst to find a way to adequately accommodate ANA econometrically.

6.4 Multinomial logit model

The first model estimated is an MNL model, which is reported in Table 6.1. Fare and time are both highly significant and of expected sign, with respondents preferring cheaper fares and shorter flights. Willingness to pay measures are split into two categories: WTP to obtain a desirable attribute level, or a one unit increase in a desirable attribute (such measures will be suffixed by $^+$ in Table 6.1 and henceforth); and WTP to avoid an undesirable attribute level, or a one unit increase in an undesirable attribute ($^-$). Willingness to pay t-ratios were calculated using the delta method. On average, respondents are willing to pay 56 cents to avoid one minute of flight time, or equivalently, \$33.50 to avoid one hour of flight time. Departure time took one of four levels in the experiment: 6am, 10am, 2pm and 6pm. These levels are dummy coded, with 6pm forming the base level. Significant parameters and WTPs are obtained for 6am and 10am, with the WTP values suggesting that respondents are, on average, willing to pay \$10.15 to depart at 6pm instead of 6am, and \$9.11 to depart at 10am instead of 6pm. The parameter and WTP for 2pm departure is not significant, suggesting an indifference between 2pm and 6pm departure, ceteris paribus.

Alternative specific constants were estimated for travel with Virgin Blue and Air NZ, with estimates being relative to travel with Qantas. An insignificant parameter for Virgin Blue suggests that, on average, respondents are indifferent to whether they fly with Qantas or Virgin Blue, ceteris paribus. There is a mean sensitivity against Air NZ however, with a WTP to avoid the airline of \$5.77. This measure captures preferences that are not accounted for via attributes in the choice experiment. It is worth noting that the same levels of fare,

Table 6.1: MNL model

	Parameter	***t*-ratio**	**WTP**	***t*-ratio**
Fare	-0.0729	-47.16	-	-
Flight time	-0.0407	-19.24	$0.56^{-}	18.94
Depart 6am	-0.7398	-9.60	$10.15^{-}	9.22
Depart 10am	0.6638	7.85	$9.11^{+}	8.11
Depart 2pm	0.0723	0.92	$0.99^{+}	0.92
Virgin Blue	0.0065	0.13	$0.09^{+}	0.13
Air NZ	-0.4201	-7.84	$5.77^{-}	7.93
Model fits				
LL(0)	-4614.17			
LL(MNL)	-2776.56			
Number of parameters, K	7			
ρ^2	0.3983			
Adjusted ρ^2	0.3972			
AIC	1.3255			
Observations	4200			
Respondents	210			

$^{+}$ WTP to obtain the attribute level, or a one unit increase in the attribute.

$^{-}$ WTP to avoid the attribute level, or a one unit increase in the attribute.

flight time and departure time were applied to all three airline alternatives. That is, no airline was presented as operating flights that tended to be cheaper or shorter than that of the competitor, or operating disproportionately at certain times of the day, as may be the case in the market. Thus, differences in the ASCs are unlikely to be the consequence of different attribute ranges across the choice alternatives. One possible influence on the ASCs is a left-to-right bias, whereby respondents are more likely to choose the first alternative of the three, which were presented side by side. The order of the alternatives was not varied, where such variation would help mitigate such a bias. While the possibility of some degree of left-to-right bias cannot be dismissed, it is believed to be minimal in this setting.

On balance, the ASCs that are associated with the airline are likely capturing, to a large extent, real life brand preferences. These in turn might be influenced by such factors as experience, marketing, and word of mouth. Although such factors remain latent, it is nonetheless worth speculating about some specific potential influences on the ASCs. At the time of the choice experiment, Air NZ did not and had never operated domestic routes in Australia under its own livery, and so a lack of experience might be impacting on Air NZ's ASC. However, Air NZ did have full ownership of the Australian airline Ansett at the time of its demise in September 2001. There was considerable ill sentiment towards the owners as a result of the collapse, and this may have remained two and a half years later, at the time of the study. There was also potentially a belief by some respondents that the Air NZ flight would have to be accessed from the international terminal, passage through which is more time consuming than the domestic terminal due to customs and more stringent security checks. Logically such hassles would exert a downward influence on Air NZ's ASC. While the MNL model demonstrates only a preference against Air NZ, more complex models presented in subsequent sections will reveal a more nuanced picture of sensitivity to each of the airlines.

Model fit statistics appear reasonable, and serve foremost as a baseline for subsequent models. Since the more complex models in this chapter all come at the cost of additional parameters, the AIC is the most appropriate statistic for evaluating relative model performance. All models estimated in this chapter utilise the 4200 observations obtained from the 210 respondents that were retained after data cleaning, and so these numbers will not be presented in subsequent tables.

6.5 Attribute nonattendance model

This section presents the results from both the IANA model, and the CANA model. Recall that the performance of the IANA model in the simulations in Section 4.4 was not encouraging. The models biased the magnitudes of the mean taste coefficients upwards, and overestimated the ANA rates. Further, there was evidence that preference heterogeneity and ANA is confounded. Nonetheless, this section tests the performance of the two models in an empirical context. Stated ANA rates for three of the attributes will provide some sense of whether the ANA rates are reasonable, and the individual specific parameter estimates will provide further insight. Initial comparisons are made between alternative specifications of correlation of ANA in the ANA model, either assuming independence (the IANA model), or allowing full correlation (the CANA model). The models also serve as a reference point for their more advanced counterparts, the RPIANA and RPCANA models, which introduce random parameters in place of a point estimate, conditional on attendance. Model fit and the reasonableness of the model outputs will be compared as the RPANA models are detailed. Despite concerns stemming from the simulations that the ANA rates from the ANA model are biased, and capturing more than just ANA, this section will, for brevity, still refer to the class constrained to zero as representing ANA.

First, Section 6.5.1 presents results from all ANA1 models, which are those that model ANA to one attribute only. Estimating the ANA1 models first allows the stability of the results to be compared as ANA is handled for all or just some of the attributes. The IANA and

CANA models are equivalent when attendance is modelled to one attribute only. Attendance to the airline is included in this analysis, even though the airline is actually presented as a choice alternative label, and modelled with an ASC. As discussed in Section 6.4, the ASC may capture other unobserved effects, such as a left-to-right bias. However, given the general prominence given to airline as a part of the product mix, it is believed that the ASC is largely capturing the mean preferences for the airlines. Consequently, a censoring of the ASCs to zero will be interpreted as nonattendance to the airline. Any reference in the text to the attendance of attributes also includes attendance to the choice alternative labels, in the interest of brevity.

Next, CANA4 and IANA4 models that handle attendance to all attributes, as well as the airline, are reported in Section 6.5.2. This allows the stability of the results to be assessed as the number of attributes for which ANA is handled increases. Also, the impact of alternative specifications of the ANA assignment models can be explored. Only the two extremes of the CANA and IANA models will be explored. 'Hybrid' specifications of the ANA assignment models will only be considered with the RPANA model. Finally, Section 6.5.3 will draw some insights from the individual specific probabilities, and discuss why using these are inappropriate.

6.5.1 Single attribute nonattendance

Four ANA1 models were estimated, handling attendance to fare, flight time, departure time, and airline. As discussed in Section 3.2, nonattendance to dummy and effects coded attributes and ASCs is best modelled by grouping over all dummy and effects coded levels and all ASCs, respectively. Specifically, if a respondent does not attend to departure time, then coefficients for all three departure time parameters (6am, 10am and 2pm) are set to zero. If a respondent does not attend to the airline, then coefficients for both ASC parameters (Virgin Blue and Air NZ) are set to zero. If not all coefficients are set to zero, then what is captured is not nonattendance, but rather an alternative expression of preference.

Table 6.2 provides an overview of the four ANA1 models, the two models handling nonattendance to all four attributes that are presented in the next section, and the MNL model and stated ANA rates, which provide reference points. All four ANA1 models have a log likelihood closer to zero than the MNL model, and have a lower AIC, despite the estimation of one extra parameter. The rates of nonattendance in the ANA1 models are detailed in the first four rows of Table 6.2. Rates of nonattendance to fare and flight time are moderate, at 13.41 and 6.5 percent respectively, with departure time and airline considerably higher at 52.74 and 92.74 percent respectively. While the accuracy of stated ANA rates can be questioned, a comparison of the stated and estimated rates is nonetheless of interest. The stated nonattendance rate of fare is somewhat lower than the estimated rate, at 6.9 percent. Conversely, the stated rate for flight time, at 18.1 percent, is somewhat higher than the estimated rate. The two rates are wildly divergent for departure time, with an estimated rate of 52.74 being far higher than the stated rate of 15.95 percent. The stated rate of nonattendance to the airline alternative labels was not collected in the survey, however, the estimated rate of 92.74 percent appears implausibly high.

Table 6.2: ANA model overview

Ignored	**Methodology**	**Classes**	**Class assign. params**	**LL**	**AIC**	**Percent ignored**			
						Fare	**F. time**	**D. time**	**Airline**
Fare	ANA1	2	1	-2636.52	1.259	13.41	-	-	-
F. time	ANA1	2	1	-2775.26	1.325	-	6.50	-	-
D. time	ANA1	2	1	-2713.70	1.296	-	-	52.74	-
Airline	ANA1	2	1	-2756.51	1.316	-	-	-	92.74
All	CANA4	16	15	-2526.75	1.214	13.63	16.33	53.42	78.98
All	IANA4	16	4	-2545.82	1.218	13.26	11.78	52.17	80.72
None	MNL	1	0	-2776.56	1.326	-	-	-	-
-	Stated ANA	-	-	-	-	6.90	18.10	15.95	-

Table 6.3 provides further detail on the ANA1 models. The first model in the table is the MNL model, for comparison. Consider the second model presented, in which nonattendance to fare can be accounted for (i.e., the ANA1 fare model). For the class that represents nonattendance to fare, the fare coefficient is set to zero, which theoretically translates to an infinite WTP. In reality, the respondent will not have a marginal disutility of money of zero, and so we can only argue that the fare was ignored over the range of fares presented (Campbell et al., 2008). Once the 13.41 percent nonattendance rate is accounted for, all WTP values that are significant in the MNL model (i.e., all except depart 2pm and Virgin Blue) decrease in magnitude, by an average of 19 percent, and ranging from a low of 16.8 percent to a high of 25.1 percent. However, the only change that is statistically significant is for depart 6am, as evidenced by the t-ratios of the differences in the WTP values between the MNL and the ANA model, reported in the column 'Diff. t-ratio'. Despite the general lack of significance, the clear trend on the WTP is downwards, which is plausible, as respondents who ignore fare will bias the fare coefficient downwards, biasing the WTP upwards, ceteris paribus. By probabilistically assigning respondents to a class with a zero fare coefficient, the

downward bias can to some extent be mitigated. This impact on the WTP can be offset by nonattendance to the attributes in the numerator of the WTP function, where the resultant WTP will be a function of the relative ANA rates for the numerator and the denominator. This will be discussed in Section 6.5.2, in the context of the ANA4 models, which handle nonattendance to all attributes.

Consider now the ANA1 flight time model. The model presents only a slight improvement in model fit over the MNL model compared to the ANA1 fare model, with reductions in AIC of 0.001 and 0.0662 units respectively. In contrast also to the ANA1 fare model is the pattern of change in WTP. While WTP for a reduction in flight time increases by 7.1 percent, other WTP values only vary slightly. This appears plausible, as only attendance to flight time is handled by the model. Once those who do not attend to flight time are accounted for, the WTP for a reduction in flight time increases. However, none of these differences in WTP are statistically significant.

One of the most notable features of the ANA1 departure time model is the strong statistical significance of the 2pm departure parameter, which was not significant in the MNL model. The positive sign of the coefficient that could tentatively be interpreted from the MNL model is confirmed, with a mean WTP to travel at 2pm instead of 6pm of \$7.74. The other key feature of the model is the statistical significance of the difference in WTP of two of the three departure time parameters, with all three exhibiting greater magnitude than in the MNL model[2]. This model presents an empirical context in which is realised the potentially inflating effect that nonattendance to an attribute has, ceteris paribus, on its WTP. The improvement in model fit compared to the MNL model is, like the ignored fare model, quite marked.

Finally, the ANA1 airline model also demonstrates an inflation in magnitude of the WTP for the attribute for which nonattendance is modelled. The Virgin Blue constant is significant in the airline model, where it was not in the MNL model. Willingness to pay to avoid flying with Air NZ is \$43.25, up from \$5.77 in the MNL model, where this difference is dramatic in its magnitude and highly significant, with a t-ratio of 20.53. Differences in WTP are not significant for all other attributes.

[2]The insignificance of the 2pm departure parameter in the MNL model must be noted however. The sign of this parameter can only be accepted with a low level of confidence.

Table 6.3: ANA models accommodating ANA for one attribute

Ignored	None (MNL)		Fare			Flight time			Departure time			Airline		
	Param.	***t*-ratio**	**Param.**	***t*-ratio**		**Param.**	***t*-ratio**		**Param.**	***t*-ratio**		**Param.**	***t*-ratio**	
Fare	-0.0729	-39.63	-0.0964	-56.72		-0.0731	-72.57		-0.0767	-66.57		-0.0735	-80.70	
Flight time	-0.0407	-18.95	-0.0447	-25.52		-0.0437	-16.06		-0.0443	-24.07		-0.0403	-21.31	
Depart 6am	-0.7398	-9.60	-0.7325	-14.28		-0.7537	-12.79		-2.0272	-16.39		-0.7969	-14.74	
Depart 10am	0.6638	7.85	0.7183	10.47		0.6579	9.27		0.8532	8.72		0.6447	9.30	
Depart noon	0.0723	0.92	0.0915	1.40		0.0637	0.98		0.5936	6.20		0.0057	0.09	
Virgin Blue	0.0065	0.13	0.0043	0.10		0.0061	0.14		-0.0752	-1.79		-0.6773	-11.33	
Air NZ	-0.4201	-7.84	-0.4623	-10.71		-0.4247	-9.72		-0.4882	-11.18		-3.1798	-16.08	
	WTP	**WTP *t*-ratio**	**WTP**	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**	**WTP**	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**	**WTP**	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**	**WTP**	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**
Flight time	$\$0.56^{-}$	18.94	$\$0.46^{-}$	24.11	0.43	$\$0.60^{-}$	16.31	0.15	$\$0.58^{-}$	22.66	0.08	$\$0.55^{-}$	20.72	0.05
Depart 6am	$\$10.15^{-}$	9.22	$\$7.60^{-}$	13.50	1.98	$\$10.31^{-}$	12.21	0.11	$\$26.41^{-}$	16.34	9.86	$\$10.84^{-}$	13.74	0.50
Depart 10am	$\$9.11^{+}$	8.11	$\$7.45^{+}$	11.03	1.24	$\$9.00^{+}$	9.83	0.08	$\$11.12^{+}$	9.13	1.31	$\$8.77^{+}$	9.89	0.24
Depart 2pm	$\$0.99^{+}$	0.92	$\$0.95^{+}$	1.41	0.03	$\$0.87^{+}$	0.98	0.09	$\$7.74^{+}$	6.31	4.44	$\$0.08^{+}$	0.09	0.66
Virgin Blue	$\$0.09^{+}$	0.13	$\$0.04^{+}$	0.10	0.04	$\$0.08^{+}$	0.14	0.01	$\$0.98^{-}$	1.78	0.97	$\$9.21^{-}$	12.22	7.81
Air NZ	$\$5.77^{-}$	7.93	$\$4.80^{-}$	11.18	0.90	$\$5.81^{-}$	9.95	0.04	$\$6.36^{-}$	11.80	0.53	$\$43.25^{-}$	16.60	20.53
	Ignored	***t*-ratio**	**Ignored**	***t*-ratio**		**Ignored**	***t*-ratio**		**Ignored**	***t*-ratio**		**Ignored**	***t*-ratio**	
Fare	-	-	13.41%	19.28[2]		-	-		-	-		-	-	
Flight time	-	-	-	-		6.50%	5.91[2]		-	-		-	-	
Departure time	-	-	-	-		-	-		52.72%	23.35[2]		-	-	
Airline	-	-	-	-		-	-		-	-		92.74%	17.87[2]	
Model fits														
LL	-2776.56		-2636.52			-2775.26			-2713.70			-2756.51		
K	7		8			8			8			8		
ρ^2	0.3983		0.4286			0.3985			0.4119			0.4026		
Adjusted ρ^2	0.3972		0.4275			0.3974			0.4108			0.4015		
AIC	1.3255		1.2593			1.3254			1.2960			1.3164		

1. t-ratio of difference between this model's WTP and MNL WTP.
2. t-ratio for difference to zero percent ANA.

$^{+}$ WTP to obtain the attribute level, or a one unit increase in the attribute.
$^{-}$ WTP to avoid the attribute level, or a one unit increase in the attribute.

6.5.2 Nonattendance to all attributes

Table 6.4 shows the results from both the CANA4 and IANA4 models, where all four attributes can be ignored. Considering first the rates of nonattendance, broadly, there is alignment between the rates recovered from the CANA4 model, the IANA4 model, and each of the ANA1 models. Nonattendance to fare is highly consistent, with rates of 13.63 percent (CANA4), 13.26 percent (IANA4) and 13.41 percent (ANA1). Nonattendance to flight time exhibits the greatest variation, with rates of 16.33 percent (CANA4), 11.78 percent (IANA4) and 6.5 percent (ANA1). This is an early flag that the assumption of independence of ANA might not hold for flight time. Departure time is very consistent between the CANA4 and IANA4 models, with rates of 53.42 (CANA4), 52.17 (IANA4) and 52.72 (ANA1) percent. Finally, ANA rates were consistent between the CANA4 and IANA4 models for airline, with rates of 78.98 percent and 80.72 percent respectively, contrasting with an ANA1 rate of 92.74 percent.

Since a single parameter controls the ANA rate for each attribute in the IANA model[3], the associated standard error can be used to provide a measure of statistical reliability of the ANA rate. Table 6.4 presents, for each IANA nonattendance parameter, a t-ratio which represents whether the ANA rate is different from zero. The difference is significant for all attributes. In contrast, the CANA model determines ANA rates by summing the class assignment probabilities of all classes that treat the attribute as ignored, and no measure of statistical confidence can be calculated at the attribute level.

The nonattendance rates can be considered, and compared between the CANA4 and IANA4 models, not just for each attribute in isolation, but for the combinations of fare and non-fare attributes. These rates are noteworthy because they reflect, for each attribute, the proportions of the sample that have a valid WTP, a zero WTP, and an infinite WTP. Table 6.5 reports the results. In the CANA4 model, the rates are the sum of the class assignment probabilities for all classes that represent the corresponding attendance patterns for the attributes in question[4]. In the IANA4 model, the rates are a multiplication of the nonattendance or attendance probabilities for fare and the WTP attribute in question. Note that nonattendance to both fare and the WTP attribute has been interpreted as a zero WTP, in contrast to the interpretation by Rose et al. (2005) of an infinite WTP. Whilst technically this ratio is infinite, behaviorally a zero WTP has more appeal, as the respondent will have

[3]So long as no covariates are introduced.

[4]Since attendance is modelled for four attributes, and only two attributes are used for the WTP measure, the rates are summed over four classes (i.e., all attendance combinations of the remaining two attributes).

Table 6.4: ANA models accommodating ANA for all attributes

Methodology	**CANA4**			**IANA4**		
	Param.	***t*-ratio**		**Param.**	***t*-ratio**	
Fare	-0.1084	-49.78		-0.1059	-53.78	
Flight time	-0.0606	-19.03		-0.0568	-19.05	
Depart 6am	-2.4617	-16.16		-2.3533	-17.72	
Depart 10am	0.9870	6.35		0.9347	7.75	
Depart noon	0.6691	4.23		0.6213	4.98	
Virgin Blue	-0.6165	-4.85		-0.6855	-6.54	
Air NZ	-2.0201	-11.37		-2.0892	-13.98	
	WTP	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**	**WTP**	**WTP *t*-ratio**	**Diff.[1] *t*-ratio**
Flight time	0.56^-$	20.68	0.00	0.54^-$	21.74	0.09
Depart 6am	22.72^-$	16.10	7.92	22.23^-$	18.20	7.92
Depart 10am	9.11^+$	6.64	0.00	8.83^+$	7.92	0.19
Depart noon	6.17^+$	4.34	3.28	5.87^+$	4.99	3.25
Virgin Blue	5.69^-$	4.79	4.25	6.47^-$	6.50	5.09
Air NZ	18.64^-$	11.38	8.37	19.73^-$	13.83	9.52
	Ignored	***t*-ratio**		**Ignored**	***t*-ratio[2]**	
Fare	13.63%	-		13.26%	26.20	
Flight time	16.33%	-		11.78%	11.81	
Departure time	53.42%	-		52.17%	40.15	
Airline	78.98%	-		80.72%	17.51	
Model fits						
LL	-2526.75			-2545.82		
K	22			11		
ρ^2	0.4524			0.4483		
Adjusted ρ^2	0.4495			0.4468		
AIC	1.2140			1.2175		

1. t-ratio of difference between this model's WTP and MNL WTP.

2. t-ratio for difference to zero percent ANA.

$^+$ WTP to obtain the attribute level, or a one unit increase in the attribute.

$^-$ WTP to avoid the attribute level, or a one unit increase in the attribute.

ignored the attribute irrespective of their nonattendance to fare. Further, the nonattendance to a cost attribute is likely to be because of inappropriate attribute level ranges.

Table 6.5: ANA model ANA rates for each combination of fare and non-fare attributes

Fare **WTP attribute** **WTP interpretation**		**Ignore** **Ignore** **Zero**	**Attend** **Ignore** **Zero**	**Ignore** **Attend** **Infinite**	**Attend** **Attend** **Valid**
Flight time	CANA4	9.51%	6.83%	4.12%	79.55%
	IANA4	1.56%	10.22%	11.70%	76.52%
Departure time	CANA4	10.19%	43.24%	3.44%	43.14%
	IANA4	6.92%	45.25%	6.34%	41.49%
Airline	CANA4	9.46%	69.52%	4.17%	16.85%
	IANA4	10.70%	70.02%	2.56%	16.72%

The rates with which respondents attend to both the fare and non-fare attributes are somewhat consistent between the CANA4 and IANA4 models, for all three WTP attributes. Considering the other three attendance combinations, low to mild differences can be observed between the CANA4 and IANA4 models for departure time and airline. More severe are the differences for flight time. Most notably, with the CANA4 model, 9.51 percent of respondents ignore both flight time and fare, while a mere 1.56 percent do so under the IANA4 model. The consistent rates of dual attendance to the WTP attributes are reassuring, as these are the percentages of the sample for which the WTP is valid. More problematic are the divergent rates for other combinations of attendance, for these rates represent the diametrically opposed WTP outcomes of a zero WTP and an infinite WTP[5]. Consider for example WTP for flight time. Under the CANA4 model, 16.33 percent of respondents have a zero WTP, while WTP is infinite for only 4.12 percent of respondents. Under the IANA4 model, 11.78 percent of respondents have a zero WTP, and compared to the CANA4 model a much larger percent of respondents have an infinite WTP: 11.70 percent. One possible cause of this divergence is the assumption of the IANA model that ANA is independent across attributes. This is a further flag that this assumption may be violated for flight time in the IANA model. This possibility will be explored extensively with the RPANA model, in Sections 6.7.4 and 6.7.5.

Next, consider the differences in WTP between the CANA4, IANA4 and MNL models. Although not reported in the tables, no WTP values differ between the CANA4 and IANA4 models to a statistically significant extent. A comparison of the CANA4 and IANA4 WTP values with their respective MNL WTP values can be made in terms of direction, magnitude, and statistical significance. The direction of change reveals which force dominates - the

[5]Or more likely, a WTP that cannot be recovered with the fare ranges presented.

downwards pull on WTP once account is made for nonattendance to fare, or the upwards pull on WTP once account is made for nonattendance to the attribute for which the WTP is being calculated. The t-ratio of the difference in WTP is reported in Table 6.4 in column 'Diff. t-ratio'.

Willingness to pay to reduce flight time is not significantly different between the MNL, CANA4 and IANA4 models. For departure time, a significant difference in WTP can be observed between both of the ANA4 models and the MNL model, for 6am and 2pm departure WTP values, but not for 10am departure, with the 6am and 2pm WTPs in the ANA models being of considerably greater magnitude. While noting an only marginally significant WTP for Virgin Blue under the MNL model, all airline WTP values in the ANA4 models differ significantly from their MNL counterparts. Again, the WTP values are of greater magnitude once attendance has been accounted for, and the differences are in many cases marked. For example, WTP to avoid flying with Air NZ increases from $5.77 with the MNL model to $18.64 and $19.73 for the CANA4 and IANA4 models, respectively.

Comparing model fits, both models offer a significant improvement on the MNL model, with drastically lower AIC values. The CANA4 model has a better log likelihood than the IANA4 model, and despite costing an additional 11 parameters, outperforms the IANA4 model, with a lower AIC value. One possible cause for this might be the aforementioned assumption of independence of nonattendance in the IANA model. The major contribution of this work is the introduction of random parameters into a generalised ANA model framework[6], resulting in the RPANA model. An outperformance of the IANA model by the CANA model does not imply that the RPCANA model will outperform the RPIANA model, even on the same dataset. It is plausible, for example, that the parsimony of the IANA assignment model structure will be a greater advantage when parameters also need to be estimated to represent taste heterogeneity. Section 6.7 will extensively explore alternative RPANA model specifications.

6.5.3 Individual specific parameter estimates

The above analysis, and the majority of the literature handling ANA through an LC model, has only considered the unconditional class assignment probabilities. However, these unconditional probabilities can be conditioned upon each individual's observed sequence of choices, to

[6] That is, the IANA and CANA models, and a range of different possible specifications allows varying degrees of independence of ANA.

obtain individual specific class assignment probabilities and parameter estimates[7] (Kamakura and Russell, 1989; Greene and Hensher, 2003). Using these individual specific estimates to make inferences about each individual's ANA behaviour is not very robust, in part because the conditional probabilities are a function of the unconditional probabilities, and problems in the latter are likely to influence the former. For example, the simulation results from Chapter 4 suggest that the ANA classes are capturing not just ANA, but weak preferences as well, and so the conditional estimates may be biased. Nonetheless, an investigation of the individual specific estimates may shed further light on the appropriateness or otherwise of the ANA model.

If a respondent is assigned with high probability to a class in which the taste coefficient(s) of an attribute is (are) constrained to zero, then it is more likely that the respondent is ignoring the attribute. Conversely, if the probability of assignment to that class is not high, then the individual is less likely to be ignoring the attribute, and their sensitivity to the attribute may be represented by some discrete mixture of coefficients across classes. It is not clear what probability value is high enough to classify the individual as not attending, or even if an extremely high probability is sufficient. The appropriate probability is uncertain, arbitrary, and not statistically informed. Certainly, the threshold of 50 percent employed by Scarpa et al. (2011, see Section 2.2.2 for details) is so low as to render meaningless any interpretation of such a classification as representing ANA. Instead, in this exploratory analysis, two tentative thresholds, 95 and 99 percent, will be tested. The number of individuals categorised with these thresholds is used to calculate an overall ANA rate informed by the individual specific estimates, which is then compared with the unconditional ANA rate.

Define the probability of assignment of individual n to class m, conditional on their sequence of choices $\vec{i}$, as $P_{m|n\vec{i}}$. Only ANA1 models will be estimated in this section, and so there will only be two classes. The class of interest is the one which represents ANA, in which the taste coefficient(s) of the attribute is (are) constrained to zero. Then, a probability threshold is specified, such that if $P_{m|n\vec{i}}$ exceeds this threshold, that individual is tentatively classified as not attending to the attribute. If the threshold is not exceeded, the attribute is classified as being attended to by the individual, and will be represented by some discrete mixture of the coefficients in each class.

Table 6.6 presents the results with the conditional class assignment probabilities. For each of the four ANA1 models, the percent of the 210 respondents with conditional assignment probabilities that exceed the threshold is reported in rows 'Zero'. The same threshold is

[7] Strictly, the conditional probabilities are associated with the sequence of choices, not the individual.

also used to assign some respondents to the second class with the estimated coefficient(s) for the attribute (rows 'Estimated'). All respondents that fail to be placed in one of these two categories are considered to have a discrete mix of coefficients (rows 'Discrete mix').

Table 6.6: ANA model conditional parameter estimate summary

Parameter(s)	**Coefficient(s)**	**Conditional probability, percent of sample where $P_{m\|n\tilde{i}}$**		**Unconditional probability**	**Stated ANA**
		>95 percent	**>99 percent**		
Fare	Zero	10.48%	9.05%	13.41%	6.90%
	Estimated	80.95%	79.05%	86.59%	-
	Discrete mix	8.57%	11.90%	-	-
Flight time	Zero	0.00%	0.00%	6.50%	18.10%
	Estimated	71.43%	28.10%	93.50%	-
	Discrete mix	28.57%	71.90%	-	-
Departure time	Zero	31.90%	18.57%	52.74%	15.95%
	Estimated	25.24%	17.62%	47.26%	-
	Discrete mix	42.86%	63.81%	-	-
Airline	Zero	86.67%	81.43%	92.74%	-
	Estimated	4.76%	3.81%	7.26%	-
	Discrete mix	8.57%	14.76%	-	-

The ANA1 fare model sees a decrease in ANA rate from 13.41 percent under the unconditional probabilities, to 10.48 percent under the conditional probabilities with a threshold of 95 percent, to 9.05 percent with a 99 percent threshold. This suggests that the class with a taste coefficient fixed to zero is capturing not just attribute nonattenders, but those with a low sensitivity to fare, consistent with the findings from Chapter 4. Encouragingly, the two conditional rates are closer to the stated ANA rate of 6.9 percent. The conditional rates for flight time, however, are less encouraging. No respondents can be classified as nonattenders under either threshold, calling into question the interpretation of the class as representing ANA.

Conditional ANA rates for departure time, at 31.90 and 18.57 percent for 95 and 99 percent thresholds respectively, are lower than the unconditional ANA rate of 52.74 percent. The second rate of 18.57 percent is quite close to the stated rate of 15.95 percent. Attribute nonattendance to airline sees a slight drop, from a suspiciously high unconditional rate of 92.74 percent to still high rates of 86.67 and 81.43 percent. The results from the RPL model in the next section will show a variety of different preference ranks for the departure times and airlines. The risk with a two class model, where one class has all parameters associated with a dummy or effects coded attribute set to zero, is that the zero, 'ANA' class serves as the best approximation of the less popular preference ranks across the attribute levels, and so consequently the rate of ANA is overestimated.

Overall, the findings with the individual specific class assignment probabilities of the ANA model are not encouraging. The discrepancy between the ANA rates generated from the conditional and unconditional probabilities, and the failure to detect any ANA for flight time, based on the conditionals, is further evidence of the inappropriateness of the ANA model. Given this, inferring individual specific ANA from the conditional probabilities of the ANA model is not recommended.

6.6 Random parameters logit model

This section presents the results from a series of RPL models. These models serve a number of purposes. First, they provide a reference point for the RPANA models, in terms of model fit and model outputs. The RPANA model can be considered in a number of ways. The most intuitive is as an extension of the RPL model. There exists a distribution of preferences across decision makers for each attribute, captured via random parameters. Additionally, there is the potential for an elevated mass at zero, captured via latent classes with some classes having taste coefficients constrained to zero, reflecting nonattendance to or indifference for the attribute by some decision makers. Thus, the latent class structure adds another layer to capture the specific behavioural motivation of ANA. If the ANA rate is not significantly different to zero, then the RPANA model will collapse to a RPL model, which is nested within it (see Section 3.3.1). Therefore, for any given RPANA model, the equivalent RPL model with the same continuous distribution for each parameter serves as a logical point of comparison. Second, the RPL model with certain distributions may result in implausibly signed coefficients. Chapter 4 demonstrated that this outcome may result from an elevated rate of true ANA, and so of interest in this empirical setting is the interrelation of implausibly signed coefficients under the RPL model, and the recovered ANA rate under the RPANA model. Obviously such a comparison necessitates estimation of RPL models. Third, the censored normal distribution can be estimated and compared to the nonattendance rates recovered with the RPANA model.

This section will document and explore numerous RPL models, with the distribution of each attribute varied one at a time. This is a natural process when trying to find the best fitting RPL model, but documentation of such an enumeration is not always warranted. The advantage here is that it allows comparison with the corresponding RPANA model, in which attendance to the varied attribute is modelled. By varying one attribute at a time, the impact of modelling ANA can be considered in isolation.

For estimation, 5000 Halton draws were employed. The motivation for such a large number of draws was discussed in Section 3.3.2. The random parameters were estimated as univariate distributions in both the RPL and RPANA models. That is, no correlation between the random parameters was modelled. Such correlation could only be introduced for normal distributions, or transformations thereof, such as the lognormal distribution, by implementing a Choleski decomposition. However, this body of work seeks to compare the model performance of a wide range of RP distributions in both the RPL and RPANA models, and univariate distributions provide this flexibility. Allowing for correlation in the continuous parameter distributions in the RPANA model will remain an area for future research.

In all models presented until this point, departure time and airline (represented, albeit not exclusively, by the ASCs) have been dummy coded. However, it has been found in this body of work that identification issues emerge when using the RPANA model to estimate ANA on dummy coded attributes, where random parameters are employed for each of the attribute levels, with distributions that are unconstrained in sign (i.e., can span zero). The problem can be overcome by employing effects rather than dummy coding, as was discussed in Section 3.3.2.

It was found that preference heterogeneity exists for departure time and airline such that the coefficients over the sampled population may span zero. Empirical tests found that zero bounded distributions, including the censored normal, constrained triangular, lognormal, Rayleigh and constrained uniform, gave poorer model fits than unbounded distributions, including the normal, triangular and uniform. Consequently, sign constraints should not be imposed upon the departure time and ASC parameters, and they will need to be effects coded when estimating the RPANA models, to overcome the identification problem noted above. To aid model comparison, they will also be effects coded for the RPL models. The best model fit was achieved with the normal distribution for both departure time and the ASCs, and this specification will be employed for all random parameter models.

Fare and flight time are attributes for which we can plausibly claim that larger values should not lead to higher utility[8]. At most, we might expect a respondent to be indifferent to the attribute, at least across the range of attribute levels presented in the context of the tradeoffs provided. Despite this, both distributions with forced sign and those without were estimated. The latter is due to the widespread use of such distributions in similar contexts, in practice, and the recognition that a mass of attribute nonattenders might lead

[8]Short haul domestic flights are unlikely to be Giffen goods.

to sign violations (see Section 4.3), producing a model that is superior statistically if inferior behaviourally.

Table 6.7 presents a summary of the RPL results with the distribution for fare and flight time varied one at a time. Specifically, as the fare distribution varied, flight time was always estimated with the normal distribution[9]. As the flight time distribution varied, the fare was estimated with the lognormal distribution[10]. Model performance, measured by the AIC, varies considerably across distributions for fare, but less so for flight time. A comparison of the different distributions will be made by drawing upon tables containing a more extensive range of model outputs.

Table 6.7: RPL model specification search

	Fare			**Flight time**		
	LL	**AIC**	**pos. or 0**	**LL**	**AIC**	**pos. or 0**
Normal	-2315.10	1.1091	3.68%	-2307.35	1.1054	3.92%
Censored normal	-2310.88	1.1071	4.77%	-2306.27	1.1049	4.96%
Triangular	-2312.54	1.1079	2.55%	-2306.78	1.1051	3.49%
Constrained triangular	-2327.44	1.1145	0%	-2310.91	1.1066	0%
Lognormal	-2307.35	1.1054	0%	-2309.42	1.1064	0%
Rayleigh	-2309.57	1.1060	0%	-2307.10	1.1048	0%
Uniform	-2305.22	1.1044	0%	-2306.30	1.1049	0%
Constrained uniform	-2314.30	1.1082	0%	-2308.42	1.1054	0%

The first set of models, which systematically vary the distribution for fare, are presented in detail in Tables 6.8 and 6.9. The models constitute a very significant improvement in model fit over the MNL model. All parameters are of expected sign[11], however, the departure time and airline parameters are not easy to interpret, due to the effects coding.

To assist in the interpretation of the random parameters for departure time and airline, Table 6.10 presents the incidence rates with which the coefficient of one effects coded attribute level exceeds (or is exceeded by) the coefficient of each other attribute level. This was achieved with a Monte Carlo simulation. Specifically, for each of the 16 models in Tables 6.8, 6.9, 6.11 and 6.12, 5000 coefficients were independently generated for each random parameter associated with an effects coded attribute, using Halton draws to improve simulation efficiency. The coefficients of the base levels were generated implicitly. Then, for

[9]The normal distribution for flight time was not in fact the best distribution, but the reduction in fit from the best distribution was not pronounced.

[10]The best distribution for fare was in fact the uniform, however, the lognormal was selected as it was found that the uniform could not be used with the RPANA models.

[11]Note that for the lognormal distribution, the attribute was multiplied by minus one, to allow estimation of a distribution in the positive domain.

Table 6.8: RPL models with alternative distributions for fare, part one

		Normal		**Censored normal**		**Triangular**		**Constrained triangular**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-0.1470	-21.74	-0.1486	-19.77	-0.1496	-20.53	-0.1508	-25.53
(varies)	σ	0.0821	14.89	0.0891	13.01	0.1933	15.14	-	-
	pos. or 0	3.68%		4.77%		2.55%		0%	
Time	μ	-0.0714	-16.29	-0.0720	-16.23	-0.0714	-16.35	-0.0684	-17.21
(normal)	σ	0.0379	8.17	0.0390	8.55	0.0380	8.10	0.0341	8.01
	pos. or 0	2.98%		3.25%		3%		2.24%	
Depart 6am	μ	-1.4350	-12.28	-1.4398	-11.92	-1.4429	-11.96	-1.3737	-12.40
(normal)	σ	1.5010	15.84	1.5050	15.78	1.4967	15.58	1.4601	14.92
Depart 10am	μ	0.9680	10.29	0.9767	10.27	0.9683	10.16	0.9233	10.73
(normal)	σ	0.9333	8.39	0.9145	8.22	0.9340	8.29	0.8668	8.01
Depart 2pm	μ	0.1059	1.16	0.1255	1.48	0.1097	1.32	0.1109	1.36
(normal)	σ	0.9879	10.56	0.9796	11.11	0.9885	12.34	0.9466	9.46
Virgin Blue	μ	0.1697	3.67	0.1694	3.69	0.1709	3.74	0.1720	3.72
(normal)	σ	0.3238	6.29	0.3383	6.45	0.3276	6.33	0.3434	6.71
Air NZ	μ	-0.4483	-8.73	-0.4498	-8.77	-0.4490	-8.64	-0.4264	-8.11
(normal)	σ	0.4251	9.14	0.4291	8.96	0.4293	8.96	0.4287	9.19
Model fits									
LL		-2315.10		-2310.88		-2312.54		-2327.44	
K		14		14		14		13	
ρ^2		0.4983		0.4992		0.4988		0.4956	
Adjusted ρ^2		0.4966		0.4975		0.4971		0.4940	
AIC		1.1091		1.1071		1.1079		1.1145	

Table 6.9: RPL models with alternative distributions for fare, part two

		Lognormal		**Rayleigh**		**Uniform**		**Constrained uniform**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0250	-30.92	-0.1229	-21.10	-0.1586	-18.93	0.1568	-18.38
(varies)	σ	0.8618	13.56	-	-	0.1493	16.85	-	-
	pos. or 0	0%		0%		0%		0%	
Time	μ	-0.0717	-15.73	-0.0715	-16.52	-0.0731	-16.60	-0.0711	-16.12
(normal)	σ	0.0407	8.19	0.0375	8.37	0.0396	8.61	0.0393	8.39
	pos. or 0	3.92%		2.83%		3.26%		3.55%	
Depart 6am	μ	-1.3466	-11.30	-1.4549	-12.27	-1.4585	-12.02	-1.3428	-11.61
(normal)	σ	1.4825	14.98	1.5127	16.45	1.5137	15.91	1.4687	13.93
Depart 10am	μ	0.9778	10.51	0.9944	10.51	0.9728	10.31	0.9691	10.08
(normal)	σ	0.8369	7.69	0.9261	8.40	0.9107	8.05	0.8266	7.46
Depart 2pm	μ	0.1410	1.59	0.1481	1.72	0.1288	1.49	0.1260	1.50
(normal)	σ	0.9761	9.64	1.0036	11.25	0.9980	11.21	0.9671	9.40
Virgin Blue	μ	0.1829	3.85	0.1733	3.77	0.1718	3.74	0.1802	3.86
(normal)	σ	0.3458	6.48	0.3459	6.66	0.3410	6.55	0.3376	6.33
Air NZ	μ	-0.4524	-8.44	-0.4522	-8.57	-0.4531	-8.72	-0.4584	-8.79
(normal)	σ	0.4372	9.08	0.4522	9.36	0.4391	8.86	0.4291	8.96
Model fits									
LL		-2307.35		-2309.57		-2305.22		-2314.30	
K		14		13		14		13	
ρ^2		0.4999		0.4995		0.5004		0.4984	
Adjusted ρ^2		0.4983		0.4979		0.4987		0.4969	
AIC		1.1054		1.1060		1.1044		1.1082	

each of departure time and airline, all combinations of attribute levels (or alternatives, for airline) were considered, and the incidence rate with which the coefficient of one of these levels (ASCs) exceeded the other coefficient was calculated[12]. Finally, each of these incidence rates were averaged across all of the models in which the distribution of fare and flight time were varied. This averaging was performed in the interest of brevity, and was supported by the consistency over the models of the structural parameters of the departure time and ASC random parameters. Rather than comparing pairs of attribute levels (or ASCs), incidence rates of each total ordering of attribute levels (ASCs) could also have been calculated (e.g., what percentage of draws represent a preference order from most to least preferred of Qantas, then Virgin Blue, then Air NZ). However, there are six total orderings for airline and 24 for departure time, and so once again for the sake of brevity, they will not be reported.

Table 6.10: Coefficient orderings for departure time and airline

D. time 1	**preferred to**	**D. time 2**	**Rate**
6am		10am	8.68%
6am		2pm	19.91%
6am		6pm	30.63%
10am		2pm	74.12%
10am		6pm	60.96%
2pm		6pm	47.78%

Airline 1	**preferred to**	**Airline 2**	**Rate**
Qantas		Virgin Blue	54.61%
Qantas		Air NZ	78.17%
Virgin Blue		Air NZ	87.43%

Drawing on Table 6.10, it can be seen that for only 8.68 percent of the draws will the 6am departure time coefficient be larger than the 10am coefficient, suggesting that most respondents prefer 10am to 6am departures[13]. A similar but slightly less pronounced effect can be observed when comparing 6am to 2pm and 6pm departures. A 10am departure provides more utility than a 2pm or 6pm departure, for 74.12 and 60.96 percent of respondents, respectively. Finally, a change in departure time from 2pm to 6pm will bring an increase in utility for about half of respondents, and a decrease for the remaining half. Slightly more than half

[12]These incidence rates were very stable with 5000 draws.

[13]These are unconditional probabilities, and so for any one individual we can only say they exhibit a certain preference structure up to a probability. For example, any one individual will prefer 6am to 10am departures with a probability of 8.68 percent. However, if we consider the sample as a whole, we can make inferences about certain percentages of that sample, for example, 8.68 percent of the sample prefer 6am to 10am departures.

of respondents prefer Qantas to Virgin Blue. Considerably more respondents (78.17 percent) prefer Qantas to Air NZ, while more again prefer Virgin Blue to Air NZ (87.43 percent). These rates appear plausible.

Of the alternative distributions for the fare parameter, the uniform results in the best model fit, with an AIC of 1.1044, followed by the lognormal distribution, at 1.1054. Notably, the imposition of constraints on the structural parameters for fare results in a worse model fit than if the same distribution remains unconstrained. For the constrained uniform distribution, the spread parameter, σ, is constrained to equal the mean parameter, μ, resulting in a uniform distribution bounded by zero and 2μ. Even though the unconstrained uniform distribution has σ (0.1493) of nearly the same magnitude as μ (-0.1586), constraining the two to equality results in a worsening of log likelihood from -2305.22 to -2314.30. The constrained triangular distribution similarly has its spread parameter constrained to equal its mean. Model fit with the constrained triangular distribution for fare is again considerably worse than for the unconstrained distribution, with log likelihoods of -2327.44 and -2312.54 respectively.

The unconstrained triangular distribution for fare has a spread that exceeds the magnitude of the mean. Consequently, 2.55 percent of coefficients are positive, which is behaviourally implausible. Nonetheless, such a distribution results in considerably better model fit than the behaviourally more appealing constrained triangular distribution. The normal distribution also has coefficients of behaviourally implausible sign (3.68 percent), but unlike the models with triangular and uniform distributions, some nonzero percentage is unavoidable, due to the unbounded nature of the distribution. The uniform distribution for fare has no coefficients of implausible sign.

One possible explanation for the differing rates of implausibly signed coefficients between the triangular and uniform distributions lies in the differences of the distributions as they approach their limits. The triangular distribution tapers to zero at the two limits, which may result in little mass near zero. By contrast, the uniform distribution does not taper, and with the same mean and spread as the triangular, will have more mass near zero. If there truly is ANA, then there may be sufficient mass near zero with the uniform distribution to adequately approximate ANA. To obtain sufficient mass, the triangular distribution may need to include coefficients of implausible sign, which will also approximate ANA. See Section 4.3.3 for a discussion on this point. The poor performance of the triangular distribution in this study might also be explained by the tapering of the distribution as it approaches its upper limit of magnitude. This tapering, combined with the finite support, might do a poor job of handling outliers. Chapter 4 demonstrated that implausibly signed coefficients can be a consequence of ANA. The role of outliers was not investigated, and some combination of the two may be influencing the results for fare in this study. However, this remains speculative.

The censored normal distribution for fare represents an improvement in model fit over the uncensored normal. The implied rate of ANA of 4.77 percent is a little less than the 6.9 percent stated by respondents. If there truly is nonattendance to fare, then any advantage by capturing ANA in the censored normal distribution is outweighed by the better model fits when fare is specified with uniform, lognormal and Rayleigh distributions. This suggests that other aspects of the true, latent distribution, including the tail, might be having a strong impact on model fit, and fare may not be adequately represented by the censored normal. The confounding of ANA and preference heterogeneity in the censored normal distribution may be detrimental to the recovery of each, as the same parameters are capturing the ANA rate and the properties of the rest of the distribution. The estimation with separate parameters of an ANA point mass and a continuous specification of preference heterogeneity may help alleviate this problem.

Tables 6.11 and 6.12 detail the models with alternative distributions for flight time. The differences in model performance are less pronounced than with alternative fare distributions. While use of the censored normal distribution leads to the best log likelihood, followed very closely by the uniform, both are outperformed by the Rayleigh in terms of the AIC, since the latter requires one less parameter. As with fare, imposing a constraint on the uniform and triangular distributions to prevent implausibly signed coefficients results in worse model fits. Use of the triangular distribution for flight time results in sign violations, while it does not with the uniform; a finding that is also consistent with fare.

The censored normal implies an ANA rate of 4.96 percent; considerably lower than the stated ANA rate of 18.1 percent. However, unlike with fare, the censored normal distribution results in one of the best model fits, outperformed only by Rayleigh, if additional parameters are penalised. Even though this distribution, which recovers an ANA rate, performs well, there may still be a degree of detrimental confounding between ANA and preference heterogeneity. Consequently, the RPANA model has the potential to recover an ANA rate for flight time, while outperforming the RPL model specified with a censored normal for this attribute.

Table 6.11: RPL models with alternative distributions for flight time, part one

		Normal		**Censored normal**		**Triangular**		**Constrained triangular**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0250	-30.92	-2.0248	-31.93	-2.0245	-30.14	-2.0523	-31.84
(lognormal)	σ	0.8618	13.56	0.8717	13.38	0.8655	13.46	0.8415	13.75
Time	μ	-0.0717	-15.73	-0.0718	-17.48	-0.0722	-16.13	-0.0737	-20.23
(varies)	σ	0.0407	8.19	0.0436	7.06	0.0981	8.58	-	-
	pos. or 0	3.92%		4.96%		3.49%		0%	
Depart 6am	μ	-1.3466	-11.30	-1.3516	-11.12	-1.3496	-11.25	-1.3044	-10.75
(normal)	σ	1.4825	14.98	1.4907	14.92	1.4861	13.15	1.4371	15.52
Depart 10am	μ	0.9778	10.51	0.9836	10.46	0.9793	9.67	0.9742	10.27
(normal)	σ	0.8369	7.69	0.8416	7.84	0.8383	7.47	0.8225	7.64
Depart 2pm	μ	0.1410	1.59	0.1426	1.47	0.1422	1.65	0.1231	1.33
(normal)	σ	0.9761	9.64	0.9795	9.57	0.9762	9.34	0.9659	10.36
Virgin Blue	μ	0.1829	3.85	0.1854	3.91	0.1841	3.85	0.1776	3.82
(normal)	σ	0.3458	6.48	0.3437	6.47	0.3447	6.39	0.3369	6.22
Air NZ	μ	-0.4524	-8.44	-0.4568	-8.47	-0.4536	-8.40	-0.4465	-8.30
(normal)	σ	0.4372	9.08	0.4426	8.93	0.4399	8.88	0.4256	9.27
Model fits									
LL		-2307.35		-2306.27		-2306.78		-2310.91	
K		14		14		14		13	
ρ^2		0.4999		0.5002		0.5001		0.4992	
Adjusted ρ^2		0.4983		0.4985		0.4984		0.4976	
AIC		1.1054		1.1049		1.1051		1.1066	

Table 6.12: RPL models with alternative distributions for flight time, part two

		Lognormal		**Rayleigh**		**Uniform**		**Constrained uniform**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0385	-29.84	-2.0326	-29.94	-2.0251	-30.46	-2.0390	-33.19
(lognormal)	σ	0.8685	13.63	0.8653	13.28	0.8735	13.50	0.8684	13.87
Time	μ	-2.7562	-46.64	-0.0584	-16.94	-0.0726	-16.08	-0.0733	-17.14
(varies)	σ	0.5748	7.45	-	-	0.0690	9.18	-	-
	pos. or 0	0%		0%		0%		0%	
Depart 6am	μ	-1.3102	-11.38	-1.3504	-12.43	-1.3559	-10.93	-1.3319	-10.51
(normal)	σ	1.4531	14.96	1.4825	15.36	1.4899	13.64	1.4646	15.43
Depart 10am	μ	0.9815	10.06	0.9777	10.60	0.9789	9.88	0.9964	10.46
(normal)	σ	0.8286	7.53	0.8408	7.75	0.8407	7.55	0.8248	7.50
Depart 2pm	μ	0.1177	1.23	0.1394	1.54	0.1405	1.69	0.1203	1.26
(normal)	σ	0.9774	9.68	0.9633	9.52	0.9718	9.67	0.9795	10.17
Virgin Blue	μ	0.1853	3.90	0.1841	3.91	0.1862	3.90	0.1840	3.92
(normal)	σ	0.3378	6.14	0.3394	6.41	0.3446	6.44	0.3408	6.20
Air NZ	μ	-0.4534	-8.45	-0.4524	-8.41	-0.4563	-8.39	-0.4552	-8.43
(normal)	σ	0.4216	9.08	0.4383	8.91	0.4435	8.84	0.4321	9.31
Model fits									
LL		-2309.42		-2307.10		-2306.30		-2308.42	
K		14		13		14		13	
ρ^2		0.4995		0.5000		0.5002		0.4997	
Adjusted ρ^2		0.4978		0.4984		0.4985		0.4982	
AIC		1.1064		1.1048		1.1049		1.1054	

6.7 Random parameters attribute nonattendance model

6.7.1 Model specification and identification

As with the RPL models, 5000 Halton draws were employed, drawing from univariate distributions. The RP distributions were constrained in numerous ways, to facilitate identification. The reader is referred back to Section 3.3.2 for an extensive discussion of the identification issues faced. The lognormal, constrained triangular and constrained uniform distributions were chosen for fare and flight time, to ensure that no sign violations occurred. Problems were encountered with the constrained uniform distribution for these attributes. To ensure identification in the RPANA model, effects coding was employed in place of dummy coding for the departure time attribute and the alternative specifics constants. Again, refer to Section 3.3.2 for more details. Once more, the caveat is that the ASCs are not strictly associated with airline, and any estimation of ANA to the ASCs cannot solely be interpreted as nonattendance to airline. For example, as discussed in Section 6.4, the ASCs may in part represent a left-to-right bias, and so ANA to the ASCs may represent a lack of such bias. However, the conclusion was that the ASCs are likely capturing airline brand preferences, to a large extent, and so nonattendance to the airline will be the predominant interpretation of ANA to the ASCs.

6.7.2 Single attribute nonattendance

The first set of RPANA models presented will only model nonattendance to a single attribute. This allows a careful comparison with the RPL model specified with the same distributions for each parameter, and indeed all RPL models that vary the distribution of the attribute of interest. Drawing upon the notation introduced in Section 3.2, a single ANA assignment model is specified ($A = 1$), which handles nonattendance to a single attribute ($K_1^\star = 1$), and contains two classes ($|C_1| = 2$). Since ANA is modelled for one attribute only, no assumption need be made about the independence or correlation of ANA (i.e., whether the model should be specified as the RPIANA or RPCANA model). For the sake of brevity, these models will be referred to as RPANA1 models, and the single attribute for which nonattendance is modelled may be appended to the end (e.g., RPANA1 fare).

The base model has lognormally distributed fare, and normally distributed flight time, departure time and airline. Table 6.13 summarises the results of the RPANA1 models. For both fare and flight time, the rate of nonattendance was estimated in conjunction with three distributions: constrained triangular, lognormal and Rayleigh. All RPANA1 models retrieve ANA rates that are statistically different to zero. However, not all RPANA1 models are statistically significant improvements over their RPL equivalents, a finding that will be discussed in detail.

Table 6.13: RPANA1 model specification search

Fare		**Flight time**		**D. time**	**Airline**	**RPANA1**		**RPL**	
distrib.	**ANA**	**distrib.**	**ANA**	**ANA**[1]	**ANA**[1]	**LL**	**AIC**	**LL**	**AIC**
Const. △	3.47%	Normal	-	-	-	-2323.23	1.1130	-2327.44	1.1145
Lognorm.	2.12%	Normal	-	-	-	-2306.78	1.1056	-2307.35	1.1054
Rayleigh	1.51%	Normal	-	-	-	-2308.89	1.1061	-2309.57	1.1060
Lognorm.	-	Const. △	10.78%	-	-	-2306.40	1.1050	-2310.91	1.1066
Lognorm.	-	Lognorm.	12.96%	-	-	-2305.87	1.1052	-2309.42	1.1064
Lognorm.	-	Rayleigh	4.82%	-	-	-2306.34	1.1049	-2307.10	1.1048
Lognorm.	-	Normal	-	29.22%	-	-2286.10	1.0958	-2307.35	1.1054
Lognorm.	-	Normal	-	-	52.93%	-2298.53	1.1017	-2307.35	1.1054

1. Both departure time and airline were distributed normally for all models.
- signifies that ANA was not modelled for this attribute.

Table 6.14 details the three RPANA models that measure attendance to fare only. The ANA rates are low for all distributions tested, ranging from 1.51 to 3.47 percent. If the ANA rates were zero, then the RPANA model would collapse to the RPL model. Therefore, the most meaningful t-ratio for the attendance parameter is of the difference between the estimated ANA rate, and an ANA rate of zero percent. The differences are significant for all three models. Another useful means of assessing the ANA rates is to construct 95 percent confidence intervals around the estimated ANA rates. Across the three models, the smallest lower bound is 0.21 percent, and the largest upper bound is 12.67 percent. The stated ANA rate of 6.9 percent is higher than all three estimated rates, but comfortably within all three confidence intervals. All three ANA rates estimated with the RPANA1 model are much lower than the ANA rate of 13.41 percent estimated with the ANA1 model (c.f. Table 6.2, p.162). Model outputs from the ANA model (i.e., with fixed coefficients) must be used with caution, as has been discussed extensively. In any case, the RPANA1 models greatly outperform any of the ANA1 models.

The RPANA model with a constrained triangular distribution for fare has a log likelihood of -2323.23. The RPL model with the same distributions is nested within the RPANA model, and so a likelihood ratio test can be performed to see if the RPANA model represents a statistically significant improvement in terms of model fit. With one degree of freedom, the test statistic of 8.40 exceeds the chi-squared critical value of 3.84 at the 95 percent confidence level ($8.40; \chi^2_{1,.05} = 3.84$), and so the null hypothesis that the two models are

Table 6.14: RPANA1 models with alternative distributions for fare

		Const. triangular		Lognormal		Rayleigh	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-0.1568	-23.93	-1.9795	-33.53	0.1255	20.03
(varies)	σ	-0.0701	-18.53	0.7906	10.09	-	-
	ANA	-3.3262	6.40[1]	-3.8333	3.17[1]	-4.1778	2.67[1]
	ANA rate	3.47%		2.12%		1.51%	
	ANA 95% C.I.	1.19%	9.71%	0.32%	12.67%	0.21%	10.19%
Flight time	μ	0.0364	7.90	-0.0722	-16.25	-0.0716	-16.19
(normal)	σ	-1.4074	-11.80	0.0409	8.07	0.0382	8.23
Depart 6am	μ	1.4658	14.70	-1.3723	-10.52	-1.4548	-12.49
(normal)	σ	0.9365	9.66	1.4735	14.64	1.5293	16.10
Depart 10am	μ	0.8689	8.03	0.9734	10.29	0.9960	10.60
(normal)	σ	0.1103	1.20	0.8428	7.73	0.9269	8.36
Depart 2pm	μ	0.9641	9.90	0.1286	1.35	0.1456	1.64
(normal)	σ	0.1739	3.84	0.9768	9.51	1.0010	11.08
Virgin Blue	μ	0.3260	6.32	0.1827	3.84	0.1749	3.80
(normal)	σ	-0.4309	-8.15	0.3392	6.36	0.3439	6.62
Air NZ	μ	0.4270	8.94	-0.4501	-8.39	-0.4542	-8.57
(normal)	σ	-3.3262	6.40	0.4329	8.98	0.4533	9.40
Model fits							
LL		-2323.23		-2306.78		-2308.89	
K		14		15		14	
ρ^2		0.4965		0.5001		0.4996	
Adjusted ρ^2		0.4948		0.4983		0.4979	
AIC		1.1130		1.1056		1.1061	

1. t-ratio for difference to zero percent ANA.

equivalent can be rejected. However, not only is the RPL model with a constrained triangular distribution for fare outperformed by every other RPL model with alternative distributions for fare, so too is the RPANA model with a constrained triangular distribution for fare and nonattendance to fare modelled. That is, the RPANA model in question is outperformed by every RPL model bar the one with a constrained triangular for fare. A comparison with the RPL model with an *un*constrained triangularly distributed fare provides some insights. This RPL model comfortably outperforms the RPANA model with a constrained triangular distribution. The RPL model spans zero, with 2.55 percent of coefficients being positive. This may be reflective of a point mass at zero (see Chapter 4), or, through the symmetry of the distribution, it may allow the tail to be longer and capture more extreme sensitivities. Given the strong performance of the lognormal distribution for fare, the latter is plausible. However, when the triangular distribution is constrained to be bounded on one side at zero, implausibly signed coefficients cannot be recovered, and so the tail of the distribution cannot be lengthened through symmetry. Thus, the constrained triangular distribution may struggle with a long tail even more than the unconstrained triangular. Definite conclusions cannot be drawn, but it does hint that the behaviour of the tail of the continuous distribution may have implications for the incidence of implausibly signed coefficients in the RPL model, and the most appropriate choice of distribution for the RPANA model.

Neither the lognormal nor Rayleigh distributions for fare in the RPANA1 model represent an improvement over their RPL model counterparts. The AIC for the RPANA model with lognormal fare is 1.056, against 1.1054 for the RPL model. Employing the likelihood ratio test, the null hypothesis that the two models are equivalent cannot be rejected ($1.14; \chi^2_{1,.05} = 3.84$). With the Rayleigh distribution, the null hypothesis that the RPANA and RPL models are equivalent also cannot be rejected ($1.35; \chi^2_{1,.05} = 3.84$). The AIC for the RPANA model, at 1.1061, is also higher than for the RPL model, at 1.1060. Furthermore, using the AIC, the RPANA model with lognormally distributed fare (1.1056) is outperformed by the RPL model with uniformly distributed fare (1.1044). The RPANA model with Rayleigh distributed fare (1.1061) is also outperformed by both the RPL model with uniformly distributed fare (1.1044), and the RPL model with lognormally distributed fare (1.1054). Therefore, as currently specified, the RPANA model offers no advantage for handling ANA to fare.

The inability of the RPANA1 fare model to outperform the RPL model may stem from the properties of the fare distributions. For example, as the mean of the lognormal distribution decreases, the tail will become fatter. Consequently, the lognormal may perform well when there is a mass at or close to zero, and a long, fat tail, capturing high sensitivity to the attribute. If, however, the ANA is largely captured through the estimated discrete point mass with the RPANA model, then this may limit the ability of the lognormal to capture the high sensitivities with the long tail. In effect, a long tail would increase the mass close to zero, which would compete with the point mass at zero. Indeed, the ANA rate with lognormally distributed fare (2.12 percent) is lower than with the constrained triangular (3.47 percent), which suggests that such confounding may be occurring. It must be stressed however that the ANA rate with the constrained triangular is not necessarily the true value. Whilst once again not conclusive, this reasoning supports the hypothesis that the tail of the continuous distribution, and indeed the distribution as a whole, will have an impact on the ANA rate estimated.

With the models presented thus far, it does not appear as if the RPANA model has much to offer with respect to nonattendance to fare in this study. The problems may in part stem from what is likely a low ANA rate. The estimated rates average just 2.37 percent, and the stated rate is 6.9 percent. Coupled with only a moderate sample size of 210 respondents, there may not be enough expression of ANA to fare in the sample to allow it to be retrieved by the model. One way to potentially improve the performance of the RPANA model is to introduce covariates into the ANA assignment models, rather than relying on an average propensity across the sample to not attend to an attribute. This will be attempted in Section 6.7.3. Another approach is to test whether nonattendance to the attributes is independent across attributes. If it is not, then modelling attendance to combinations of attributes might lead to an improved model fit over the RPL model. Specifically, a RPANA model could be estimated with more than one attribute in the ANA assignment model a that handles nonattendance to the problem attribute (i.e., $K_a^* > 1$). Such an approach will be presented in Section 6.7.4.

The three RPANA models that measure attendance to flight time only are presented in Table 6.15. The ANA rates range from 4.82 percent for the Rayleigh distribution to 12.96 percent for the lognormal distribution, and so exhibit greater variation than the alternative models that handle nonattendance to fare. Each rate is significantly different to zero. The confidence intervals range from a smallest lower bound of 0.71 percent to a largest upper bound of 30.13 percent. The flight time ANA rate under the ANA model, at 6.5 percent, lies between the RPANA rates with the Rayleigh and constrained triangular distributions. However, again, all the RPANA models strongly outperform the ANA model. All three rates estimated with the RPANA model are somewhat lower than the stated ANA rate of 18.1 percent, but again the stated rate lies comfortably within all three confidence intervals.

Table 6.15: RPANA1 models with alternative distributions for flight time

		Const. triangular		**Lognormal**		**Rayleigh**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0257	-29.19	-2.0221	-27.73	-2.0223	-30.09
(lognormal)	σ	0.8586	13.57	0.8595	13.35	0.8716	13.12
Flight time	μ	-0.0833	-16.30	-2.5338	-31.51	-0.0615	-14.32
(varies)	σ	-	-	0.3897	4.95	-	-
	ANA	-2.1139	8.82[1]	-1.9043	9.22[1]	-2.9826	3.92[1]
	ANA rate	10.78%		12.96%		4.82%	
	ANA 95% C.I.	4.00%	25.94%	4.89%	30.13%	0.71%	26.54%
Depart 6am	μ	-1.3320	-10.66	-1.3374	-10.42	-1.3670	-10.85
(normal)	σ	1.4730	15.34	1.4752	15.37	1.4963	14.75
Depart 10am	μ	0.9944	10.45	0.9910	10.17	0.9845	10.36
(normal)	σ	0.8334	7.43	0.8368	7.43	0.8467	7.69
Depart 2pm	μ	0.1257	1.32	0.1208	1.25	0.1341	1.44
(normal)	σ	0.9893	10.07	0.9933	9.92	0.9722	9.54
Virgin Blue	μ	0.1829	3.81	0.1838	3.76	0.1869	3.91
(normal)	σ	0.3460	6.22	0.3489	6.39	0.3436	6.47
Air NZ	μ	-0.4568	-8.46	-0.4558	-8.32	-0.4561	-8.40
(normal)	σ	0.4360	9.03	0.4364	8.90	0.4400	8.81
Model fits							
LL		-2306.40		-2305.87		-2306.34	
K		14		15		14	
ρ^2		0.5001		0.5003		0.5002	
Adjusted ρ^2		0.4985		0.4985		0.4985	
AIC		1.1050		1.1052		1.1049	

1. t-ratio for difference to zero percent ANA.

Notably, all estimated rates for fare and flight time are lower than their stated rates. One possibility is that respondents are, on aggregate, overstating their propensity to not attend to an attribute. If this is the case, the actual rates of stated nonattenders who do attend and stated attenders who do not attend is not known, and indeed some mix of the two might lead to the estimated rate. Evidence resolving this question will be presented in Section 6.7.3. Alternatively, the RPANA model may not be recovering the true rate, possibly due to the ANA point mass capturing some portion of the continuous component of utility representing preference heterogeneity.

The RPANA1 model with a constrained triangular distribution for flight time represents a statistically significant improvement in model fit over the RPL model with the same distributions ($9.02; \chi^2_{1,.05} = 3.84$). An improvement is also observed with the lognormal distribution ($7.09; \chi^2_{1,.05} = 3.84$). However, the RPANA1 model does not lead to a statistically significant improvement in model fit when the Rayleigh distribution is used for flight time ($1.53; \chi^2_{1,.05} = 3.84$).

Comparing the RPANA and RPL models on the AIC, with the distribution for flight time varying only, the RPANA model with Rayleigh distributed flight time (1.1049) is slightly outperformed by the RPL model with the Rayleigh distribution (1.1048), and matched by the censored normal and uniform distributions. The RPANA constrained triangular model (1.1050) is outperformed by the RPL model with Rayleigh, censored normal and uniform distributions. The RPANA lognormal model (1.1052) is additionally outperformed by the RPL triangular model (1.1051). As with fare, the RPANA1 flight time model is not leading to an improvement in model fit over the RPL model. However, as was noted in the discussion of the RPANA1 fare models, two avenues will be explored to try and improve the performance: covariates in the ANA assignment models, and relaxing the assumption of independence in ANA across the attributes. Both provide encouraging results.

The first model in Table 6.16 models nonattendance to departure time only. Normally distributed departure time parameters give the best performance for both the RPL and RPANA models. For fare and flight time, the extent of sign violation under various distributions with the RPL model was of interest, together with the performance of the corresponding RPANA models. This is not of interest here, so only the best RPANA model, utilising the normal distribution, is reported. Compared to the equivalent RPL model, the RPANA model represents a large, statistically significant improvement in model fit ($42.50; \chi^2_{1,.05} = 3.84$). The ANA rate is 29.22 percent, with a confidence interval of 14.14 to 50.85 percent. The stated ANA rate of 15.95 percent lies towards the low end of this range, while the rate recovered by the

ANA model exceeds the top end of the range, at 52.74 percent. There may be confounding to some extent between the ANA rate recovered and the continuous component of departure time utility, as for some draws from the RP distributions of departure time, the coefficients of all departure times will be close to zero, and so provide a good approximation of ANA. However, unless the means of the random parameters are close to zero (they are not in this instance), these combinations of coefficients are unlikely to have a high rate of incidence. If the incidence rate is more common, empirical identification problems may result.

Table 6.16: RPANA1 models for departure time and airline

		Departure time		**Airline**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0162	-25.72	-2.0007	-30.03
(lognormal)	σ	0.8982	12.89	0.8534	13.06
Flight time	μ	-0.0758	-15.05	-0.0608	-13.08
(normal)	σ	0.0407	7.57	-1.3928	-10.30
Depart 6am	μ	-1.9953	-8.18	1.4780	14.35
(normal)	σ	2.0484	12.25	0.9959	10.38
Depart 10am	μ	1.5333	11.57	0.8365	7.63
(normal)	σ	0.9425	6.48	0.1142	1.15
Depart 2pm	μ	0.4343	2.78	1.0019	9.42
(normal)	σ	1.3703	9.25	0.3462	3.61
	ANA	-0.8846	12.85[1]	-	-
	ANA rate	29.22%		-	
	ANA 95% C.I.	14.14%	50.85%	-	-
Virgin Blue	μ	0.1551	3.28	0.5252	5.06
(normal)	σ	0.3443	6.40	-0.8839	-6.30
Air NZ	μ	-0.4643	-8.41	0.5110	6.53
(normal)	σ	0.4590	8.52	-3.2343	2.99
	ANA	-	-	0.1173	15.39[1]
	ANA rate	-		52.93%	
	ANA 95% C.I.	-	-	31.49%	73.34%
Model fits					
LL		-2286.10		-2298.53	
K		15		15	
ρ^2		0.5045		0.5019	
Adjusted ρ^2		0.5028		0.5001	
AIC		1.0958		1.1017	

1. t-ratio for difference to zero percent ANA.

The second model in Table 6.16 models nonattendance to airline only. Again, the only model reported has normally distributed airline parameters, which provide the best model fit. This model fit is a considerable improvement on the equivalent RPL model ($17.63; \chi^2_{1,.05} = 3.84$). The ANA rate is sizeable, at 52.93 percent, with a confidence interval of 31.49 to

73.34 percent. There was no stated ANA collected, with which the estimated rate can be compared. The implausible ANA rate under the ANA model of 92.74 percent exceeds the upper end of the confidence interval.

6.7.3 Covariates in attribute nonattendance

All of the RPANA models estimated thus far have treated the probability of attribute nonattendance as being the same across respondents. However, some respondents may be more or less likely to attend to an attribute, and it may be that these differing propensities can be linked to socio-demographics, attitudinal information, or any other variables that the analyst can observe. Recall that Section 2.2.3 discussed the concept of ANA heterogeneity, and the role that covariates may play in the recovery of ANA. Section 3.2 contained details on how these covariates enter into the RPANA model, and control the ANA rate at the respondent level. Obviously, the use of covariates to vary the probability of ANA across respondents is motivated by a desire to improve model fit, and gain insights into why attributes may not be attended to. More specifically in the context of this study, covariates may allow the RPANA model to outperform the RPL model for fare and flight time. Finally, covariates could be used as a way to leverage stated ANA responses.

Presumably, those who state that they ignore an attribute are more likely to ignore it than those who state otherwise. Nonetheless, stated ignorers may still attend to the attribute, and stated attenders may ignore it. The RPANA model, with stated ANA as a covariate, can accommodate these scenarios probabilistically, and so not be reliant on the very strong assumption that stated ANA is completely accurate and free from error. However, the approach may still suffer from a problem of endogeneity.

As a preliminary investigation into the accuracy or otherwise of stated ANA responses in this empirical setting, a RPL model was estimated that specified two sets of parameters for each attribute. One set of parameters was estimated for those who stated that they attended to the attribute, while another was estimated for those who stated that they ignored the attribute. This approach has been employed in numerous studies (e.g., Hess and Rose, 2007; Hess and Hensher, 2010; Campbell and Lorimer, 2009). Table 6.17 presents the results for this model, with two sets of parameters estimated for all attributes for which stated ANA was collected: fare, flight time, and departure time. To aid interpretation of the lognormal distributions, the means and medians of the actual lognormal distributions are reported, in addition to the means and standard deviations of the underlying normals. With the exception of 2pm departure time, the sensitivities to attributes for stated ignorers are estimated with significant random parameters, of expected sign. However, the mean sensitivities for stated ignorers are about half the magnitude of those for stated attenders. These findings are in line with existing studies.

Table 6.17: RPL model with separate parameter estimates based on stated ANA

		Stated attended		**Stated ignored**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-1.9577	-24.40	-2.5504	-11.93
(lognormal)	σ	0.8933	12.19	0.7285	3.83
	mean[1]	0.2104	-	0.1018	-
	median[1]	0.1412	-	0.0780	-
Flight time	μ	-0.0795	-14.97	-0.0484	-5.41
(normal)	σ	0.0410	7.02	0.0309	2.24
Depart 6am	μ	-1.5217	-9.33	-0.7178	-3.58
(normal)	σ	1.7257	13.18	0.7703	5.62
Depart 10am	μ	1.1041	10.23	0.6260	2.37
(normal)	σ	0.9244	7.88	0.5339	1.85
Depart 2pm	μ	0.1870	1.55	0.0474	0.33
(normal)	σ	1.1846	8.86	-	-
Virgin Blue[2]	μ	0.1706	3.72	0.1706	3.72
(normal)	σ	0.3149	5.92	0.3149	5.92
Air NZ[2]	μ	-0.4723	-8.90	-0.4723	-8.90
(normal)	σ	0.4335	9.16	0.4335	9.16
Model fits					
LL		-2286.41			
K		23			
ρ^2		0.5045			
Adjusted ρ^2		0.5018			
AIC		1.0997			

1. Of the lognormal distribution, not the underlying normal.
2. No stated ANA, so separate parameters not estimated.

One conclusion to draw from the above model is that at least some of the stated ignorers appear to be attending to the attribute. As with many other studies, the accuracy of stated ANA responses is called into question. It should be noted that the estimated sensitivities for stated ignorers do not imply that all or even most such respondents still attend to the attribute. Just as random parameters can mask ANA in conventional RPL models, the random parameters estimated for stated ignorers may be masking a subset of respondents who truly are ignoring the attribute, as they claim to be doing. For example, the normal distribution for flight time for stated ignorers implies that 5.84 percent of coefficients have implausible sign. These coefficients may be approximating ANA, as discussed in Chapter

4. Thus, some stated ignorers may truly be ignoring the attribute. Others may just have a muted sensitivity relative to stated attenders. Still other may have a sensitivity no different to stated attenders, especially if their stated ANA response was simply erroneous.

The model in Table 6.17 does not provide much information about nonattendance behaviour of respondents that claim to have attended to an attribute. Consider though the rate of sign violation for flight time for those who claimed to have attended to the attribute. At 2.61 percent, it is less than that of stated ignorers at 5.84 percent. This may be because the ANA rate for stated attenders is lower than than for stated ignorers, thus requiring fewer coefficients of implausible sign to approximate ANA. Whilst only weak evidence, it hints at differences in ANA between stated attenders and ignorers, and therefore at the information that may be contained in stated ANA responses, despite the presence also of error. In total, the above model suggests that stated ANA in this empirical context is of limited reliability, and should not be used to deterministically set the marginal utilities of attributes of stated ignorers to zero. The RPANA model is well placed to leverage the stated ANA information, but handle it probabilistically, whilst also capturing preference heterogeneity.

Three RPANA1 models were estimated, each modelling ANA for a single attribute, with the stated ignoring response for that attribute, for respondent n, included as a dummy in the vector of covariates z_n. The three models correspond to the three attributes for which stated ANA was captured: fare, flight time, and departure time. The fare model failed to converge, perhaps due to the estimation of an additional parameter for the covariate, when the baseline model without the covariate only yielded an estimated ANA rate of 3.47 percent.

The flight time model could be estimated, and strongly outperformed the baseline model without the covariate. Table 6.18 presents the results of this model. Fare utilises the lognormal distribution, flight time the constrained triangular, and attendance is only modelled for flight time. The baseline model, documented in Table 6.15, has a log likelihood of -2306.40. The introduction of the covariate results in a log likelihood of -2298.87, for the cost of only one more parameter, representing a significant improvement in model fit ($15.06; \chi^2_{1,.05} = 3.84$). This RPANA model, with an AIC of 1.1019, now clearly outperforms the best RPL model, which utilises the Rayleigh distribution and has an AIC of 1.1048. Nonetheless, the potential problem of endogeneity has to be recognised by the analyst, and stated ANA responses need to be collected, where the RPANA model was motivated in part by a desire to be relieved of such a burden.

The implied ANA rates for flight time for both stated attenders and stated ignorers are telling. Stated attenders have an ANA rate of 5.21 percent, suggesting that a small proportion

Table 6.18: RPANA1 models with stated ANA as a covariate for modelled nonattendance

		Flight time		**Departure time**	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0208	-34.80	-2.0167	-25.85
(lognormal)	σ	0.8651	13.55	0.8952	13.17
Flight time	μ	-0.0858	-15.70	-0.0757	-14.99
(varies)	σ	-	-	0.0403	7.56
	Distribution	Const. triangular		Normal	
	ANA constant	-2.9011	8.97	-	-
	ANA rate, stated attended	5.21%		-	
	ANA stated ignored	2.9735	10.44	-	-
	ANA rate, stated ignored	51.81%		-	
Depart 6am	μ	-1.3533	-10.64	-1.9890	-8.31
(normal)	σ	1.4771	15.37	2.0658	12.13
Depart 10am	μ	1.0030	10.35	1.5288	11.73
(normal)	σ	0.8438	7.55	0.9623	6.83
Depart 2pm	μ	0.1172	1.18	0.4489	3.14
(normal)	σ	1.0121	10.26	1.3828	9.98
	ANA constant	-	-	-1.3335	19.20
	ANA rate, stated attended	-		20.86%	
	ANA stated ignored	-	-	2.3311	13.40
	ANA rate, stated ignored	-		73.06%	
Virgin Blue	μ	0.1822	3.81	0.1564	3.33
(normal)	σ	0.3455	6.26	0.3434	6.42
Air NZ	μ	-0.4567	-8.42	-0.4653	-8.50
(normal)	σ	0.4321	8.97	0.4558	8.25
Model fits					
LL		-2298.87		-2275.93	
K		15		16	
ρ^2		0.5018		0.5068	
Adjusted ρ^2		0.5000		0.5049	
AIC		1.1019		1.0914	

of stated attenders actually ignore flight time. Of those respondents who stated that they ignored flight time, only 51.81 percent actually did so. Both of these findings support the argument that stated ANA is not reliable. That only half of stated ignorers actually ignore the attribute is particularly important, as it suggests that simply constraining to zero the partworth utility for these respondents is untenable. However, as a source of information, stated ANA can improve RPANA model fit.

Table 6.18 also documents the RPIANA model that utilises stated nonattendance to departure time as a nonattendance covariate. With a log likelihood of -2275.93 and one additional parameter, the covariate model strongly outperforms the baseline model (see Table 6.16) that has a log likelihood of -2286.10 ($20.34; \chi^2_{1,.05} = 3.84$). Stated attenders have an ANA rate of 20.86 percent, and stated nonattenders 73.06 percent. Both of these rates are higher than for flight time, implying that stated attendance responses are less accurate for departure time than for flight time, but stated nonattendance responses are more accurate.

No stated ANA responses were collected for airline. The only socio-demographic variables collected were gender and age. Introducing these as ANA covariates for each attribute in turn failed to lead to any improvement in model fit, likely due to the homogeneity of the sample. Overall, it is found that introducing stated ANA as a covariate in the ANA assignment models has the potential to lead to significant improvements in model fit. An alternative way to accommodate this would be to interact stated ANA with the standard deviation parameter of the censored normal distribution. However, just as this distribution may confound ANA and preference heterogeneity generally, so to may it specifically with the stated ANA covariate. That is, by increasing the ANA rate, the covariate might also distort the sensitivities, including those that are very low and very high

In summary, there is an improvement in model fit over the covariate-free RPANA models, when stated ANA is introduced as a covariate in the flight and departure time ANA assignment models, but not when introduced for fare. Further, for flight time, introducing these covariates overcomes the inability of the RPANA model to outperform the RPL model.

6.7.4 Correlation in attribute nonattendance for fare and flight time

Another possible reason for the lack of improvement in model fit when moving from the RPL to the covariate-free RPANA model for fare and flight time is that nonattendance to the two attributes may not be independent, and thus an assumption of the RPIANA model is violated. This assumption can be relaxed by combining fare and flight time into the one

ANA assignment model. Consequently, the shares of all *combinations* of fare and flight time nonattendance are estimated.

Table 6.19 estimates such a model, with $A = 3$ ANA assignment models, one each for handling departure time, airline, and the combination of fare and flight time. Consistent with the first covariate model from Table 6.18, fare is distributed with the lognormal distribution, flight time with the constrained triangular, and the departure times and ASCs with the normal. During estimation, the share of the combination of ignored fare and attended flight time approached zero, and so this class was dropped from the ANA assignment model (i.e., $|C_{Fare.FlightTime}|$ equals three, rather than four, as when all combinations are modelled).

Table 6.19: RPANA model with correlation in ANA for fare and flight time

		Param.	***t*-ratio**	
Fare	μ	-1.9498	-29.54	
(lognormal)	σ	0.7372	10.74	
Flight time	μ	-0.0828	-15.70	
(const. △)	σ	-	-	
Depart 6am	μ	-1.3301	-10.87	
(normal)	σ	1.4925	14.13	
Depart 10am	μ	1.0039	11.37	
(normal)	σ	0.8420	7.58	
Depart 2pm	μ	0.1251	1.42	
(normal)	σ	0.9893	11.04	
Virgin Blue	μ	0.1770	3.97	
(normal)	σ	0.3403	6.35	
Air NZ	μ	-0.4541	-8.57	
(normal)	σ	0.4284	8.01	
Fare	**F. time**	**Param.**	**s.e.**	**Rate**
Ignore	Ignore	-3.0833	0.5283	4.13%
Attend	Ignore	-2.7745	0.7517	5.63%
Attend	Attend	-	-	90.24%
Model fits				
LL	-2301.31			
K	15			
ρ^2	0.5013			
Adjusted ρ^2	0.4995			
AIC	1.1030			

Comparing this model, with 15 parameters and a log likelihood of -2301.31, with the the RPL model with the same distributions (Table 6.11, p.179, fourth model), with 13 parameters and a log likelihood of -2310.91, we can reject the null hypothesis that this model is not an improvement over the RPL model ($19.21; \chi^2_{1,.05} = 5.99$). A log likelihood ratio test cannot be performed between this model and the RPIANA2 model with the same distributions and

ANA for fare and flight time, as the two model specifications do not nest. However, on the AIC, this model (1.1030) outperforms the equivalent RPANA1 model (1.1050; Table 6.15, p.186, first model), the best fitting RPANA1 model (1.1049; Table 6.15, p.186, third model, lognormal fare and Rayleigh flight time), and the best RPL model tested (1.1044; Table 6.9, p.175, third model, uniform fare and normal flight time). Thus, relaxing the assumption of independence of ANA for fare and flight time results in a RPANA2 model that outperforms the RPL model.

The ANA rates for the fare-flight time combinations are noteworthy. First, the share of fare nonattendance combined with flight time attendance approaches zero (since $\gamma_{c_a} \to -\infty$). Since a rate can be estimated for fare nonattendance combined with flight time nonattendance, this implies that fare is only ignored when flight time is also. It is notable that the ANA rate for fare in this model, at 4.13 percent, is higher than the rate of 2.12 percent with the RPANA1 model with the same lognormal fare distribution (Table 6.14, p.183, second model). While the difference is small in absolute terms, it is more considerable in relative terms. This suggests that if the independence assumption is violated, estimation of the RPIANA model may not only worsen model fit, but bias the ANA rates as well. However, definitive conclusions cannot be drawn from this one comparison, and any comparison is undermined by a lack of knowledge of the true rate. The aggregate ANA rate for flight time in the model presented in this section is 9.76 percent, compared to 10.78 percent with the RPANA1 model (Table 6.15, p.186, first model). The difference between these two rates is notably smaller in relative terms than for fare.

6.7.5 Tests for independence of attribute nonattendance across attributes

The previous section presented a model that allowed for correlation in ANA between fare and flight time, resulting in an improvement in model fit from all models previous. Whether ANA between other attributes is independent or correlated should be determined. To this end, a series of models were estimated that modelled attendance to all pairs of attributes. For each pair, two models were estimated. The first, denoted as a RPCANA2 model, does not assume independence of ANA and estimates the incidence rates of all four combinations of ANA across the two attributes. Less combinations of ANA were estimated if during estimation the incidence rate of specific combinations tended to zero. The second, denoted as a RPIANA2 model, assumes independence, and so estimates the ANA rate of each attribute directly. The two models are compared on model fit, using the AIC, since the two RPANA specifications are not nested. Also, the implied probabilities of all four ANA combinations were calculated for the RPIANA2 models, and compared to the RPCANA2 models, to see for which combinations, if any, there is a notable discrepancy. The RPCANA2 ANA rates were also aggregated for each attribute, and compared with the rates estimated with the RPIANA2[14] models. By comparing model fit and ANA rates for these two models, we have some indication of whether the attributes are independent. There may be higher order correlations in ANA between three or more attributes, but considering pairs is a good starting point.

Table 6.20 shows the results for the two models with ANA modelled for fare and flight time. The full results are not reported, with the focus instead being on the ANA rates and model fits. The first four rows detail the incidence rates of the combinations of ANA, where for the RPIANA model these are multiplications of the appropriate attribute attendance or nonattendance probabilities. The next two rows show the aggregate ANA rates for each attribute, where for the RPCANA model these are just the sums of the rates for the appropriate ANA combinations.

Table 6.20: RPANA independence tests: fare and flight time

ANA combination		**ANA incidence rates**	
Fare	**F. time**	**RPCANA**	**RPIANA**
Ignore	Ignore	4.13%	0.12%
Attend	Ignore	5.63%	10.20%
Ignore	Attend	0.00%[1]	1.07%
Attend	Attend	90.24%	88.61%
Ignore	-	4.13%	1.19%
-	Ignore	9.76%	10.32%
Model fits			
LL		-2301.31	-2306.16
K		15	15
ρ^2		0.5013	0.5002
Adjusted ρ^2		0.4995	0.4984
AIC		1.1030	1.1053

1. ANA combination not estimated in final model.

The ANA rate approached zero for fare nonattendance and flight time attendance, and so this combination was removed from the model. The RPCANA model outperforms the RPIANA model on the AIC, with values of 1.1030 and 1.1053 respectively. This finding is also strongly supported by the Akaike likelihood ratio index test (Ben-Akiva and Swait, 1986). Only the ANA rates for dual attendance to fare and flight time align across the two

[14]The suffix of 2 will be dropped for the remainder of this section, for the sake of brevity.

models. While the RPIANA model implies a very low ANA rate of 0.12 percent for dual nonattendance, the RPCANA model has a much higher rate of 4.13 percent. A large discrepancy can also be observed for fare attendance combined with flight time nonattendance, with the RPIANA rate of 10.20 percent much higher than the RPCANA rate of 5.63 percent. The difference is low in absolute terms for nonattendance to fare and attendance to flight time, however, it is notable that the rate is zero for the RPCANA model. In aggregate, the ANA rates for flight time are similar across the models, but the RPCANA model recovers a higher ANA rate for fare. Overall, the evidence suggests that the assumption of independence of ANA for fare and flight time can be rejected.

The results for fare and departure time are detailed in Table 6.21. All combinations of ANA are retained with the RPCANA model. The RPIANA model outperforms the RPCANA model on the AIC. While there are some minor discrepancies, the ANA rates are quite similar across the two models, both in terms of the combinations, and for each attribute in aggregate. For fare and departure time, the assumption of independence of ANA appears to hold.

Table 6.21: RPANA independence tests: fare and departure time

ANA combination		**ANA incidence rates**	
Fare	**D. time**	**RPCANA**	**RPIANA**
Ignore	Ignore	1.53%	0.52%
Attend	Ignore	25.79%	27.47%
Ignore	Attend	0.79%	1.34%
Attend	Attend	71.89%	70.67%
Ignore	-	2.32%	1.86%
-	Ignore	27.32%	27.99%
Model fits			
LL		-2288.76	-2289.38
K		16	15
ρ^2		0.5040	0.5038
Adjusted ρ^2		0.5021	0.5021
AIC		1.0975	1.0973

Table 6.22 presents the results for fare and airline. Nonattendance to both attributes was dropped from the RPCANA model. The RPCANA model outperforms the RPIANA model on the AIC (1.1022 to 1.1024), although the difference is not dramatic. Applying the Akaike likelihood ratio index test, the upper bound on the probability of erroneously choosing the incorrect model over the true specification is 0.196, providing a moderate level of confidence that the RPCANA model is most appropriate. The most notable difference in ANA rate is for fare nonattendance combined with airline attendance, at 2.37 percent for

the RPCANA model, and 0.84 percent for the RPIANA model. Overall the differences are not great. Based on the AIC, the assumption of independence could be rejected, but the closeness of these values means that there may be value in testing both specifications in a more complex model.

Table 6.22: RPANA independence tests: fare and airline

ANA combination		**ANA incidence rates**	
Fare	**Airline**	**RPCANA**	**RPIANA**
Ignore	Ignore	0.00%[1]	1.06%
Attend	Ignore	56.31%	54.80%
Ignore	Attend	2.37%	0.84%
Attend	Attend	41.32%	43.30%
Ignore	-	2.37%	1.90%
-	Ignore	56.31%	55.86%
Model fits			
LL		-2299.57	-2299.94
K		15	15
ρ^2		0.5016	0.5015
Adjusted ρ^2		0.4998	0.4998
AIC		1.1022	1.1024

1. ANA combination not estimated in final model.

Table 6.23 presents the results for flight and departure time. The RPIANA model outperforms the RPCANA model on the AIC, with a log likelihood value that is nearly the same, despite requiring one less parameter. All ANA rates align very closely. It can be concluded that ANA for flight and departure time is independent.

Table 6.23: RPANA independence tests: flight time and departure time

ANA combination		**ANA incidence rates**	
F. time	**D. time**	**RPCANA**	**RPIANA**
Ignore	Ignore	3.20%	2.68%
Attend	Ignore	24.26%	24.89%
Ignore	Attend	6.46%	7.03%
Attend	Attend	66.07%	65.40%
Ignore	-	9.67%	9.71%
-	Ignore	27.47%	27.57%
Model fits			
LL		-2286.43	-2286.47
K		16	15
ρ^2		0.5045	0.5045
Adjusted ρ^2		0.5026	0.5027
AIC		1.0964	1.0959

Table 6.24 presents the results for flight time and airline. The RPIANA model slightly outperforms the the RPCANA model on the AIC. There is a reasonable degree of discrepancy between the incidence rates of the combinations of ANA, although the difference is less marked when comparing the overall ANA rates for each attribute. In sum, it can tentatively be concluded that the ANA for flight time and airline is independent, but less clearly so than for some other pairs of attributes.

Table 6.24: RPANA independence tests: flight time and airline

ANA combination		**ANA incidence rates**	
F. time	**Airline**	**RPCANA**	**RPIANA**
Ignore	Ignore	1.08%	4.69%
Attend	Ignore	51.89%	45.20%
Ignore	Attend	8.07%	4.71%
Attend	Attend	38.96%	45.39%
Ignore	-	9.15%	9.41%
-	Ignore	52.97%	49.89%
Model fits			
LL		-2297.63	-2298.42
K		16	15
ρ^2		0.5021	0.5019
Adjusted ρ^2		0.5001	0.5001
AIC		1.1017	1.1016

Finally, Table 6.25 shows that ANA appears to be independent for departure time and airline, with the AIC of the RPIANA model lower than for the RPCANA model. There is also a close alignment of the ANA rates, for each combination and in the aggregate.

Table 6.25: RPANA independence tests: departure time and airline

ANA combination		**ANA incidence rates**	
F. time	**Airline**	**RPCANA**	**RPIANA**
Ignore	Ignore	13.93%	12.74%
Attend	Ignore	32.56%	34.13%
Ignore	Attend	13.54%	14.44%
Attend	Attend	39.97%	38.69%
Ignore	-	27.47%	27.19%
-	Ignore	46.48%	46.87%
Model fits			
LL		-2283.74	-2283.79
K		16	15
ρ^2		0.5051	0.5050
Adjusted ρ^2		0.5032	0.5033
AIC		1.0951	1.0947

Overall, the ANA independence tests reveal that ANA appears to be largely independent across attributes. The standout exception is between fare and flight time. Allowing for correlation in ANA between these attributes leads to improved model fit, and different inferences about ANA, including a higher ANA rate for fare. There may be some mild correlation between ANA to fare and airline, based primarily on the AIC of the two RPANA models. The remaining attributes appear to have independence in ANA. Again, a caveat should be added, that there may be higher order correlations between three or more attributes. However, the tests introduced here may be useful for the analyst as a part of the model specification search.

6.7.6 Hybrid random parameters attribute nonattendance model

This section presents a single RPANA model in which nonattendance to all attributes is modelled. Informed by the results from the previous section, the model treats ANA to fare and flight time as correlated, but assumes that ANA over all remaining attributes is independent, including between the remaining attributes and the combination of fare and flight time. Consequently, it is neither the extremes of an RPIANA nor RPCANA model, and will instead be referred to as a hybrid RPANA model. As with the models reported in Tables 6.19 and 6.20 (p.194 and p.196), the combination of fare nonattendance and flight time attendance is omitted from the model on empirical grounds. For consistency with previous models presented, fare is specified with the lognormal distribution, and flight time with the constrained triangular[15].

The current model nests the two RPANA1 models that were presented in Table 6.16 (p.188), as well as the model from Table 6.19 (p.194) that handled correlated ANA for fare and flight time. Log likelihood ratio tests reveal that all three previous models are outperformed by the current model. In addition to reporting the ANA rate estimated with the current model, the table contains an extra column, 'Rate (RPANA1/2)', reporting the rates that were estimated in the two RPANA1 models (for departure time and airline), and the RPANA model that handled correlated ANA for fare and flight time. A comparison of the rates reveals a fairly high level of consistency, with the notably exception of airline, in which the current model estimates an ANA rate of 44.3 percent, and the RPANA1 model a rate of 52.93 percent. The potential to once again introduce stated ANA into this model as covariates for the ANA assignment models will remain unrealised, but would be trivial.

[15]The Rayleigh distribution performs very slightly better on the AIC for flight time in the RPANA1 models reported in Section 6.7.2, but the difference in model fit is negligible.

Table 6.26: Hybrid RPANA model

		Param.	t-ratio				
Fare	μ	-1.9112	-29.21				
(lognormal)	σ	0.7600	9.77				
Flight time	μ	-0.0834	-14.64				
(const. $\triangle$)							
Depart 6am	μ	-1.9702	-9.73				
(normal)	σ	1.9985	12.39				
Depart 10am	μ	1.4931	9.48				
(normal)	σ	0.9475	6.16				
Depart 2pm	μ	0.3643	2.69				
(normal)	σ	1.3563	10.20				
Virgin Blue	μ	0.2730	3.64				
(normal)	σ	0.4666	6.41				
Air NZ	μ	-0.7926	-5.92				
(normal)	σ	0.5074	5.59				
Fare	**F. time**	**D. time**	**Airline**	**Param.**	**s.e.**	**Rate**	**Rate (RPANA1/2)**
Ignore	Ignore	$\star$	$\star$	-3.1174	0.5041	4.05%	4.13%
Attend	Ignore	$\star$	$\star$	-3.0116	0.7031	4.50%	5.63%
Attend	Attend	$\star$	$\star$	-	-	91.45%	90.24%
$\star$	$\star$	Ignore	$\star$	-0.9805	0.2614	27.28%	29.22%
$\star$	$\star$	Attend	$\star$	-	-	72.72%	70.78%
$\star$	$\star$	$\star$	Ignore	-0.2290	0.5705	44.30%	52.93%
$\star$	$\star$	$\star$	Attend	-	-	55.70%	47.07%
Model fits							
LL	-2275.35						
K	17						
ρ^2	0.5069						
Adjusted ρ^2	0.5049						
AIC	1.0916						

$\star$ signifies independence of ANA between starred and non-starred attributes.

6.7.7 Random parameters correlated attribute nonattendance model

The final model presented is an RPCANA model that handles nonattendance to all four attributes. That is, it utilises a single ANA assignment model, and allows any degree of correlation in ANA across attributes to be captured. Motivation for such a model comes from the possibility that ANA is not independent across *any* attributes, and that failure to capture such correlation will likely be detrimental to model fit and the model outputs. In this dataset, evidence presented so far suggests that this is not the case. Nonetheless, it is worth investigating what issues are faced, and what results are obtained, when an RPCANA model is specified when independence of ANA may hold between only some attributes.

The same RP distributions are employed as in the previous section. Initially, all 16 ANA combinations were modelled, requiring 15 parameters in the ANA assignment model. However, it was apparent that not all combinations could be supported. Classes were removed in a stepwise fashion. The most obvious problem in the first model estimated lay in the four classes representing fare nonattendance and flight time attendance. The log likelihood became flat, and the standard errors for the ANA assignment parameters associated with these combinations became extremely large. This is consistent with the the models estimated in Sections 6.7.4, 6.7.5 and 6.7.6, where the incidence rate of this ANA combination was found to be zero. Consequently, these four combinations were dropped from the model specification. Three more ANA combinations were dropped, because their incidence rate approached zero. In order of their removal, these combinations were:

1. Fare nonattendance, flight time nonattendance, departure time attendance, and airline nonattendance;
2. Fare attendance, flight time nonattendance, departure time nonattendance, and airline nonattendance; and
3. Fare attendance, flight time nonattendance, departure time attendance, and airline nonattendance.

The final specification modelled nine combinations of ANA.

Table 6.27 details the model results. While the log likelihood is better than the model in the previous section that makes some ANA independence assumptions (-2272.11 verses -2275.35), it comes at a cost of four additional parameters. A log likelihood ratio test cannot be performed since the models do not nest, but the RPCANA model presented here is inferior on the AIC (1.0920 verses 1.0916).

The first column for ANA rates, 'Rate', reports the ANA rates as estimated with this

Table 6.27: RPCANA model

		Param.	***t*-ratio**				
Fare	μ	-1.8917	-26.58				
(lognormal)	σ	0.7293	9.86				
Flight time	μ	-0.0840	-15.05				
(const. $\triangle$)							
Depart 6am	μ	-1.9548	-7.91				
(normal)	σ	1.9923	12.42				
Depart 10am	μ	1.4997	10.58				
(normal)	σ	0.9466	6.59				
Depart 2pm	μ	0.3439	2.25				
(normal)	σ	1.3587	8.75				
Virgin Blue	μ	0.2957	3.03				
(normal)	σ	0.4699	4.88				
Air NZ	μ	-0.8432	-6.28				
(normal)	σ	0.5218	6.19				
Fare	**F. time**	**D. time**	**Airline**	**Param.**	**s.e.**	**Rate**	**Rate**[1] **(Hybrid)**
Ignore	Ignore	Ignore	Ignore	-3.0596	0.8919	1.60%	0.49%
Attend	Attend	Ignore	Ignore	-1.0069	0.4836	12.49%	11.05%
Attend	Attend	Attend	Ignore	0.0190	0.4677	34.83%	29.46%
Ignore	Ignore	Ignore	Attend	-3.2006	0.9016	1.39%	0.62%
Attend	Ignore	Ignore	Attend	-3.0610	0.9609	1.60%	0.68%
Attend	Attend	Ignore	Attend	-1.3173	0.6390	9.15%	13.90%
Ignore	Ignore	Attend	Attend	-2.8500	0.7140	1.98%	1.64%
Attend	Ignore	Attend	Attend	-2.5082	1.6017	2.78%	1.82%
Attend	Attend	Attend	Attend	-	-	34.18%	37.04%
Ignore	-	-	-	-	-	4.97%	4.05%
Attend	-	-	-	-	-	95.03%	95.95%
-	Ignore	-	-	-	-	9.36%	8.55%
-	Attend	-	-	-	-	90.64%	91.45%
-	-	Ignore	-	-	-	26.24%	27.28%
-	-	Attend	-	-	-	73.76%	72.72%
-	-	-	Ignore	-	-	48.92%	44.30%
-	-	-	Attend	-	-	51.08%	55.70%
Model fits							
LL	-2272.11						
K	21						
ρ^2	0.5076						
Adjusted ρ^2	0.5051						
AIC	1.0920						

1. Rates do not sum to 100%, as some some ANA combinations from the hybrid model are not reported here.

model. The rates above the dashed line are for each ANA combination, and sum to 100 percent. The rates below the dashed line sum the appropriate estimated rates to obtain the total attendance and nonattendance rates for each attribute. The next column, 'Rate (Hybrid)', reports the ANA rates obtained in the hybrid RPANA model estimated in the previous section. The rates for each ANA combination, above the dashed line, can be inferred by multiplying the appropriate attendance or nonattendance probability for each attribute (or combination of attributes, for fare and flight time). However, since some combinations were dropped in the model reported here, the reported percentages do not sum to 100.

Comparing between the two models the incidence rates for nonattendance to each combination of ANA, there is a broad alignment, with some moderate differences for some combinations. Comparing the total ANA rates for each attribute, discrepancies are only evident for airline. This may be due to the tentative conclusion drawn in Section 6.7.5 that ANA may not be independent between fare and airline. Interestingly however, the model introduced in this section estimates an ANA rate for airline (48.92 percent) that is back closer to the rate from the RPANA1 model (52.93 percent, see Table 6.16), up from a low of 44.3 percent in the hybrid model from the previous section.

In sum, the hybrid RPANA model from Section 6.7.6 is probably more appealing than the RPCANA model presented here. Model fit is slightly better on the AIC, and the model is more parsimonious. Further, the hybrid model allows ANA covariates such as stated ANA to be entered more directly against the attribute itself.

6.8 Discussion

In summary, the RPANA model outperformed the ANA and RPL models in the empirical application in this chapter. The model fit was vastly better than the ANA model, and the ANA rates were lower, and more plausible. Model fit was improved over the RPL model, however, care was required in the handling of correlation of ANA in the model. The ANA and RPANA models both suggested that the assumption of independence of ANA could not be sustained for fare and flight time, and failure to account for this in either model was detrimental to model performance. Use of stated ANA as a covariate improved model fit, and also provided some insight into the nature of the stated ANA responses. An extensive discussion of the performance of the RPANA model will be provided in the next chapter, in Section 7.1. The discussion will consider the empirical evidence from this chapter, and the evidence from the simulations in Chapter 5. The discussion will also extend to the broader implications of these findings.

The findings of this chapter also inform us about ANA behaviour in the context of choice of short haul flights. As expected, nonattendance to fare occurred with very low frequency. Nonattendance to flight time was more prevalent, but still infrequent. Heterogeneous preferences were evident for departure time, with only a minority preferring early morning flights. However, a moderate share of respondents were indifferent to departure time. More insight could have been gained into departure time if the context had been more clearly outlined to the respondent. For example, the optimal departure time for a flight from Sydney to Melbourne in the context of a weekend trip is likely to be late afternoon or early evening on the Friday. Earlier on Friday, or Saturday morning, would be less favourable. Approximately half of respondents ignored airline. This demonstrates a lack of brand loyalty by many, although this may have been somewhat reflective of the convenience sample of students. A worthy extension would be the study of ANA in the context of choice of long haul flights. Such flights will have more attributes, which may include the number of stops, stop location, code share arrangements, in-flight entertainment (IFE) options, and more legroom variability. These extra attributes are also reflective of a greater diversity of market offerings in terms of quality. With more attributes, cost-benefit influences on attribute attendance are likely to be more prevalent. Also, with more attributes, it is more likely that some of them will be ignored due to a genuine indifference to the attribute.

Chapter 7

Discussion and conclusion

This final chapter opens with a critical evaluation of the performance of the RPANA model. Then, some broader issues with ANA are discussed, and some areas for future research suggested. Finally, the key findings of the thesis are summarised.

7.1 Performance of the random parameters attribute nonattendance model

One of the most striking findings from the empirical study in Chapter 6 was the impact of the assumption of independence of ANA. For fare and flight time, relaxation of the assumption resulted in incidence rates for each of the four combinations of ANA that were very different to any that could be achieved with the independence assumption in place. Model fit improved significantly. Thus, the correct specification of the ANA assignment models appears to be crucial to the performance of the RPANA model. This is not to say that the ability of the model to leverage independence of ANA through a more parsimonious model specification is not useful. This parsimony is an advantage, so long as the data supports the assumption of independence. There is a risk involved, though. A violation of the assumption is unlikely to prevent a RPANA model relying on the assumption from being estimated. However, model fit may be compromised, and the ANA rates and other model outputs may be biased. It is advised that checks be performed. Comparing the consistency of the ANA rates for all pairs of attributes is one way to achieve this. Section 6.5.2 did this with the ANA model for all combinations of fare and non-fare attributes[1]. This may be a suitable way to gain initial insight into possible correlation, but more reliable results will be obtained with the RPANA

[1] That is, all pairs of attributes that can be used for WTP calculations.

model, as in Section 6.7.5. Alternatively, the two extremes of correlation and independence could be tested over all attributes (Hess et al., 2011). The problem here is that estimating the RPCANA model over all attributes may be slow, and unstable, due to very low incidence rates for some combinations of ANA. Indeed, Hess et al. (2011) reported stability problems with the RPCANA model, which they abandoned for the RPIANA model. They also reported some decrease in model fit when moving to the RPIANA model, which suggests that the independence assumption may not have held in all cases. Use of a hybrid RPANA model, which can rely on the independence assumption for some combinations of attributes only, may have been more appropriate than the models they estimated. These findings provide a clear answer to the question posed by Hess et al. (2011) as to whether the independence assumption is justified broadly. The assumption is context dependent.

The RPANA model clearly outperformed the ANA model in the empirical setting. This is unsurprising, as it is also capturing preference heterogeneity directly, without resorting to some discrete mix of zero and an estimated coefficient (recall Section 6.5.3). Notably, the ANA rates were lower and more plausible with the RPANA model than the ANA model, which is consistent with the findings from the simulations of Chapter 4 that the ANA rates are biased upwards by the ANA model. The combined findings of the simulations and the empirical study suggest that the ANA model is biased and cannot sufficiently separate ANA and preference heterogeneity. Use of this model is not recommended, except perhaps as an early diagnostic tool prior to the estimation of the RPANA model.

In the empirical study, the RPANA model outperformed the RPL model. Every attribute contributed to this improvement, by handling ANA for that attribute, although this was contingent upon correlation in ANA being appropriately handled. The RPL models with censored normal distributions are particularly important reference points, as they can handle ANA using conventional choice modelling methodology. These models were also outperformed by the RPANA model for fare and flight time, based on model fit. The censored normal distribution was not appropriate for departure time or airline, as random parameters for these attributes legitimately spanned zero, representing different preference orderings for the various departure times and airlines. To some extent, ANA was approximated by the coefficients near zero for these attributes under the normal distribution in the RPL model. However, this approach has limited flexibility and is difficult to interpret. The RPANA model can separate out an elevated mass at zero, allowing the structural parameters of the RP distributions to principally capture taste heterogeneity, not ANA. The RPANA model also allowed a covariate, stated ANA, to be entered against ANA specifically, rather than, for example, the structural parameters of the censored normal distributions, which handle both ANA and taste heterogeneity. In summary, the RPANA model was a clear improvement upon the RPL model with censored normal distributions. Despite this, the censored normal distribution should not be ruled out completely, as it may perform better in certain circumstances, and also could be a useful diagnostic tool for detecting ANA in a RPL model, with a RPANA model subsequently estimated for comparison if required.

As with the simulations, the RPL model appeared to be approximating ANA in the empirical application, through mass near zero, including coefficients of implausible sign. There was also some evidence to suggest that the recovery of ANA with the RPANA model, or approximations of ANA with the RPL model, may have been influenced by other parts of the distribution, in particular the tail. Unconstrained distributions may have been approximating both ANA and very large sensitivities. For example, the RPANA1 fare model with a constrained triangular distribution for fare had a poorer model fit than a RPL model with an unconstrained triangular distribution for fare, in which there was some degree of sign violation. The sign violation may have simultaneously approximated ANA, and very high sensitivities, by lengthening the tail of the distribution, through its symmetry. To support this reasoning, the RPANA1 fare model with lognormally distributed fare also outperformed the RPANA1 fare model with the constrained triangular distribution. Here, in addition to the RPANA model capturing ANA through the point mass at zero, the high sensitivities could be better handled through the tail of the lognormal distribution. The conclusion to draw is that a number of distributions should be tested with the RPANA model. In particular, symmetric and asymmetric distributions may each have their place, with asymmetric distributions particularly suited to handling very high sensitivities. However, it may be worth investigating whether some other decision rule, such as lexicographic choice, is leading to the very high sensitivities. If so, an alternative mechanism for handling these responses may be more appropriate.

The above discussion about the possible interaction of ANA and high sensitivities raises questions about whether the continuous and discrete components of the distributions can be completely separated in the RPANA model. For example, the ANA rates tended to vary across the tested distributions, with the lowest for the Rayleigh distribution, followed by the lognormal and constrained triangular distributions. The Rayleigh and lognormal distributions allow relatively more mass to be placed near zero than the constrained triangular, due to their asymmetry, and this mass may in part still be approximating ANA[2]. Alternatively, with the

[2]Much as the RPL model can approximate ANA through mass near zero.

constrained triangular distribution, the ANA point mass may still be approximating some low sensitivities[3]. The confounding of ANA and low sensitivities has certainly been reduced with the RPANA model from the ANA model. However, some confounding may still remain, as noted by Hess et al. (2011). Note that Hess et al. (2011) only estimated the lognormal distribution. Testing more distributions, such as the constrained triangular and Rayleigh, is recommended, as employing these other distributions in the RPANA model may better approximate the mix of ANA and tastes across the sample. This would likely be reflected in the model fit. Employing other distributions that are constrained in sign may further improve the accuracy with which ANA and taste heterogeneity are separately identified. In particular, more flexible distributions, such as Johnson's S_B (Johnson, 1949), might better fit the data. Also, a rich set of covariates for ANA may help disentangle ANA and low sensitivities[4]. Ultimately, we cannot be certain that ANA and low sensitivities have been separated completely. The greater the number of alternative specifications tested, and the greater the complexity of the available specifications, the more confidence we can place in the final specification adequately representing that which is latent.

The RPANA model exhibits a number of advantages over the use of stated ANA. It does not require stated ANA responses to be collected, and utilises observed choices, not stated behaviour. The unreliability of stated ANA responses can to some extent be accommodated by estimating separate coefficients for stated attenders and stated ignorers. However, the RPANA model can offer more than this approach, by using stated ANA responses as covariates in the ANA assignment models, as was demonstrated in Section 6.7.3. This alternative is consistent with the stated ANA responses exhibiting error, and allows the influence of stated ANA responses to be estimated, rather than taken as a given. Crucially, in comparison to the estimation of separate coefficients, the RPANA approach measures ANA from the stated responses *and* from information extracted via the panel nature of the responses. Other covariates may be entered into the RPANA model as well, adding flexibility.

An important practical consideration of the RPANA model is the difficulty of the specification search. Decisions need to be made about both the distributions of the random parameters[5], and the assumption of independence of ANA (and consequently the specification of the ANA assignment component of the RPANA model). Whilst constraining the sign of the

[3]As was shown with the ANA model in Chapter 4.

[4]It may be advisable to enter covariates into the taste sensitivities as well, lest the ANA covariates capture drivers of low sensitivity, exacerbating the problem.

[5]Certainly in terms of which distribution, but for the constrained triangular, alternative constraints can be tested.

coefficients in the RPANA model is a necessary condition (unless effects coding is applied), it is not sufficient. For example, problems were encountered with the constrained uniform distribution. Thus, the threat of identification problems arising during the specification search is present. Finally, the slow estimation times resulting from the computational complexity of the model means that the specification search is time consuming. The estimation times are exponential with respect to $K^{\star}$, the number of attributes for which ANA is modelled. This problem may be abated somewhat by the current trend towards multi-core processors, and utility computing (Armbrust et al., 2010), both of which allow multiple specifications to be estimated in parallel.

In addition to slowing the specification search, the computational complexity of the model may make it outright infeasible, if the number of attributes is very large. Ironically, in many situations, as the number of attributes increases, so will the incidence of ANA. The parametric expense of the RPANA model may also be great for a large number of attributes, although this will depend on the extent to which ANA is independent across the attributes. The more attributes for which independence holds, the less the parametric expense. The RPANA model by construction relies on ANA being consistent across the length of the panel, and so is likely to primarily detect serial ANA. Serial ANA may be of most interest to the analyst, as it is less context dependent, but this does not mean that choice task ANA is unimportant and unworthy of consideration. Serial and choice task ANA will be discussed further in the next section.

The various models for handling ANA considered herein were primarily evaluated in terms of model fit, and differences in parameter estimates and WTP. In part, this is a consequence of the importance placed on WTP in several areas of the literature that have considered ANA, most notably environmental economics. The models could have been evaluated in a number of further ways. Rather than consider the non-zero WTPs for those that attended to an attribute, the WTPs could have been calculated as a weighted average of zero and non-zero WTPs. For welfare assessment, the weighted average would be the most appropriate measure, and the comparison with a WTP that takes no account of ANA would be most relevant. For marketers, the WTPs for each combination of attendance and nonattendance would be useful, to assist with market segmentation[6]. Another useful way to compare models that account for ANA and those that do not is to examine differences in market share predictions. Of interest is the magnitude of any difference in shares, and whether they are substantial. Specific changes to policy or planning could be evaluated with and without ANA, to see if

[6] I wish to thank an examiner for this point.

the modelling of ANA has some impact. Finally, it would be useful to perform out-of-sample predictive validity tests, rather than solely evaluating the models within sample.

It must be acknowledged that the RPANA model was tested on only one empirical dataset in this body of work. Nonetheless, the model performed well with that dataset, and various insights were gained into the characteristics and nuances of the RPANA model. The use of the model with other datasets may provide further insights, but will be reserved for future research. Testing the model on a dataset with a large number of attributes would be valuable, despite the computational burden. The advantage of the independence assumption would be more apparent, as the RPCANA model, with no independence assumption, would become parametrically burdensome. However, as demonstrated in Chapter 6, this assumption may not hold, and so the feasibility of any RPANA model with a large value of $K^{\star}$ is uncertain.

7.2 General discussion and future research

The models estimated in Chapter 6 all relied on independently generated RP distributions. The covariances of the RP distributions were not estimated, in part because these are only meaningful with normal distributions or transformations of the normal. Correlated draws could be tested with the RPANA model, most meaningfully with the lognormal distribution. Such a test would allow an investigation of whether correlation in the discrete components of the model (i.e., ANA) is to some extent representing correlation in the continuous components (i.e., the continuous distribution, representing preference heterogeneity amongst attribute attenders). A derivation of the conditional parameter estimates would be useful. This is complicated by the RPANA model having both discrete and continuous components, as well as the ability to specify multiple ANA assignment models. Unlike the ANA model though, the unconditional estimates can handle both ANA and taste heterogeneity, and so they are not likely to significantly bias the conditional estimates. Derivation of the choice elasticities of the RPANA model would provide a more complete set of model outputs. All of the above are obvious directions for further research.

This thesis has demonstrated some of the dangers stemming from the continuous distributions employed in random parameter choice models. The simulations of Chapter 4 demonstrated that the RP distributions struggle to cope with point masses of any sizeable magnitude. Point masses at zero have a biasing influence on the various moments and descriptive measures of the continuous distribution of utility, and may induce or exaggerate the incidence of implausibly signed coefficients. Employing a distribution which enforces sign will overcome this second problem, but in doing so contribute to the first. Whilst not examined in this thesis, most commonly employed RP distributions are likely to struggle to some extent when the true preference distribution has multiple nonzero modes.

Rose et al. (2012b) demonstrated, with simulations, that lexicographic behaviour has an upwards biasing influence on the mean and standard deviation of a normal distribution, and called into question more broadly what standard deviation parameters might actually be capturing. Hess et al. (2012) drew the same conclusions from an empirical study, observing a very large reduction in random taste heterogeneity after retrieving only a small incidence rate of lexicographic choice. Given this upwards bias, it is plausible also that lexicographic behaviour might interact with ANA in terms of their influence on the RP distribution. For example, the two behaviours might interact such that downward bias in the mean from ANA cancels out the upward bias in the mean from lexicographic choice. The measure of dispersion would likely still be biased upwards. Alternatively, lexicographic behaviour might bias the standard deviation of a censored normal distribution upwards, exaggerating the size of the point mass at zero, and the ANA rate.

These problems with RP distributions stem to a large extent from the lack of flexibility that the distributions afford. Most RP distributions used in the literature are controlled by only one or two parameters, and thus there is a limit to the amount of complexity that the distribution can capture. One exception is the Johnson S_B distribution, which uses four parameters and can generate a number of fundamentally different shapes, including ones with two modes (Train and Sonnier, 2005). However, Train and Sonnier (2005) cautioned that estimation of all four parameters can suffer from identification problems. Section 2.2.2 detailed the mixtures of distributions approach, with a particular focus on their application to ANA. For any given random parameter, the structural parameters of an arbitrary number of distributions are estimated, as are parameters controlling the discrete mix of the distributions. Multiple modes can readily be accommodated, and flexibility can be increased by adding more distributions (Fosgerau and Hess, 2008), although data requirements are likely to be higher. Nonetheless, use of the mixtures of distributions approach to accommodate ANA has not been encouraging (Fosgerau and Hess, 2008; Campbell et al., 2010a, and refer to Section 2.2.2), with ANA usually captured only as low sensitivity to the attribute. In the RPANA model, this thesis has proposed a more flexible distribution, which in addition to preference heterogeneity between individuals that attend to an attribute, can handle the behaviourally plausible phenomenon of ANA.

The problem with employing inflexible distributions is that if the true preference heterogeneity is more complex than the distribution, incorrect inferences can be drawn from the model. Much of the mass of the distribution may merely be a consequence of the inflexibility of the distribution, rather than reflecting true sensitivities. Rose et al. (2012b) questioned what the standard deviation of a normal distribution may in fact be representing, given the biasing influence on the parameter of lexicographic choice. This thesis has shown that much of the mass is serving as an approximation of ANA. Questions must therefore be raised about the validity of the model outputs. Particularly concerning is WTP. Many of the WTP values may only be an artefact of the estimation technique, rather than being behaviourally grounded. To some extent, conditional parameter estimates may help, but despite the conditioning, any bias in the unconditional distributions is likely to bias the conditional distributions as well.

Ultimately, any RP distribution will merely be an approximation of the truth. The analyst will test a number of distributional forms, and select the one that best fits the data. Even the best distribution tested may be a poor approximation of the truth, although it may be difficult or even impossible to discern this. Testing more distributions, and more flexible distributions, will increase the confidence that a good fit has been achieved. To this end, the RPANA model expands the toolkit available to the analyst.

Given the problems noted above with mixed logit models with continuous mixing distributions (i.e., the RPL model), it may be more appealing to employ a discrete mixing distribution instead. The most common incarnation is the LC discrete choice model (Kamakura and Russell, 1989), although other forms will be discussed below. No assumptions need to be made about the distributions of the random parameters. This thesis has demonstrated through simulation that constraining some coefficients to zero in the LC model, in an attempt to represent ANA, is problematic. Instead of ANA being captured exclusively, some of the taste heterogeneity is captured as well. This confounding is reduced once sufficient flexibility is added into the model, as was achieved with the RPANA model through the use of random parameters. More broadly, this demonstrates that the imposition of constraints on LC models may not lead to the recovery of what the analyst wishes to recover.

Alternatively, ANA could be recovered in the LC model through freely estimated coefficients in each class. Swait and Adamowicz (2001) implemented a variant of the conventional LC model, and interpreted insignificant parameters as the associated attributes being ignored. They only identified two classes, resulting in a coarse snapshot of attendance behaviour. Nonetheless, the classes were behaviourally appealing, with one representing high levels of attribute attendance, and the other a more brand driven choice. One problem with LC models is that large numbers of classes can be parametrically expensive, and difficult to estimate classically. Train (2008) proposed using the Expectation-Maximisation (EM) algorithm (Dempster et al., 1977) as a way to overcome this problem, and successfully estimated an LC model with 20 classes.

Another approach is the discrete mixture (DM) model (Gopinath, 1995; Hess et al., 2007b). Whereas with the LC model shares are estimated for combinations of coefficients, the DM model estimates shares for each coefficient for each attribute independently. This implies that the nonparametric distribution for each attribute is independent across attributes, in a similar vein to the IANA model. In contrast, the LC model captures correlation across the taste coefficients, through the combinations of coefficients in each class, bearing similarities to the CANA model. To handle ANA, each attribute could be zero, either through free estimation, or the imposition of a constraint. However, the assumption of independently distributed tastes may be too strong.

Drawing upon the work of Bajari et al. (2007), Train (2008) estimated a highly flexible model, wherein coefficients were fixed for each attribute, and the shares estimated. This approach requires a very large number of parameters to be estimated, although again, Train (2008) achieved this with the EM algorithm. He found evidence of ANA, through an elevated mass of coefficients at zero. Of concern, however, was instability in the results as the fixed coefficients were varied across model runs. Indeed, the specification of the range of coefficients for each attribute is an analyst input. In terms of complexity, the choice of this input lies somewhere between the straightforward choice of the number of classes in the LC model, and the choice of distributions in the RPL model. As with the RPL model, the choice of this input may have a strong impact on model outputs. Unlike the RPL model, not much is known at this stage about the impact of alternative inputs. This technique, and its ability to handle ANA, shows much promise, but more research is required.

An alternative way to handle ANA is to estimate a choice model for every individual. Those attributes that are insignificant could be interpreted as not attended to. However, it may not be clear whether the insignificance stems from genuine ANA, or an insufficient number of choice responses. Indeed, Gilbride et al. (2006) noted that such an approach is likely to be unfeasible, due to the heavy data requirements, with a large number of observations required per individual. They contrasted individual level models with RP distributions, in which they noted that "information is borrowed across respondents" (p.421). They introduced the stochastic attribute selection model, which is grounded in the Bayesian framework, and combines information that is shared across respondents with the specific choices of each individual. Bayesian posteriors are similar to the conditional parameter estimates in a classically estimated RPL model. Indeed, Train (2009) showed that the two are equivalent. Thus, Bayesian posteriors and conditional parameter estimates both provide individual level estimates, without resorting to the estimation of separate models for each individual. Some effort has been made in recent years to estimate individual specific models, despite the challenges. For example, Louviere et al. (2008) estimated individual specific models by combining specific experimental design techniques with the elicitation of best-worst responses rather than most-preferred responses. This technique shows promise, but work remains to demonstrate its applicability over a range of choice contexts.

One intriguing avenue for collecting an adequate amount of information for the estimation of individual specific models is using revealed preference data sourced from online interactions. Not only may choice information be collected, but search and other process information as well. Whilst online interactions are performed over many different sites by each individual, and repeat purchases on single sites may be for very different types of goods and services (consider Amazon, for example), there is also potential to collect, from a single source, repeated choices by an individual for the one product class. For example, an individual might repeatedly use an airline website or online travel agent[7] to book flights; a rail travel booking site[8] to book train tickets; or a trip planner[9] to plan public transport trips. From the multiple choices per individual, there is an opportunity not just to improve the precision of conditional parameter estimates, or estimate individual specific models, but to identify ANA as well. However, these choices would be spread over time, and consequently, preferences, decision rules and engagement may vary along the panel, perhaps due to different choice contexts or trip purposes. For example, preferences and even attribute attendance may differ based on whether the travel is for work or recreation. Wi-Fi internet access on a train may be crucial for work travel, but ignored for recreational travel. Such differences over time would pose a challenge to the analyst seeking to exploit the panel nature of the data. The discussion will return now to SC studies, and in particular the experimental designs that such studies require.

Numerous studies have suggested that ANA might result from inappropriate attribute level ranges in SC experimental designs (Alemu et al., 2011; Hensher et al., 2012a). Closely related is the criticism that individuals should always be sensitive to cost, so long as the cost is sufficiently high (Scarpa et al., 2009; Balcombe et al., 2011; Carlsson et al., 2010). Indeed,

[7]For example, Expedia, Zuji or Travelocity.

[8]For example, thetrainline.com or MyTrainTicket.

[9]For example, www.131500.com.au.

this criticism could arguably be extended to any continuous attribute, such as travel time, the battery life of an electronic device, or the proximity of a fishing location to the nearest access point. The attribute level could be increased in magnitude until it is used by the respondent when trading between the alternatives. This is in contrast to categorical attributes such as the presence of IFE on a plane, or the requirement to purchase specific types of feed, in the context of pig species choice (Scarpa et al., 2003). There are natural limits to the states that such attributes can take, and none of these states may warrant attendance.

As a counterpoint, whilst raising the level of a continuous attribute to a magnitude that will induce trading will likely be possible, doing so might result in implausible choice scenarios, which might undermine the credibility of the study in the eyes of the respondent. The risk then is that the respondent does not take the choice tasks seriously, to the detriment of the quality of the choice responses. For example, in a route choice SC study, a high income individual may be highly insensitive to tolls. A toll of \$25 may be required for them to switch alternatives, but they may question why such tolls are even being proposed, not consider the study to be a serious undertaking, and allocate their high income time accordingly. Also, there may be political fallout, if for example it becomes known in the wider community that \$25 tolls are being proposed for a new toll road. Just because an attribute could assume levels that would ensure full attendance across a sample does not mean that it is a good idea. It may be better to present realistic attribute level ranges[10], handle ANA within the model, then recognise when interpreting the model that some individuals will not attend to an attribute across any plausible levels. In conclusion, the level of ANA should be appropriate in the context of the choice tasks presented. Whilst it is not desirable to have ANA induced by an inappropriate experimental design, it should also not be eradicated if the cost is the credibility of the choice tasks or even the entire study.

Attribute nonattendance might have implications for the concept of dominance in SC tasks. Dominance is a situation whereby one alternative is equal to or better than another alternative, on all attributes, and better on at least one attribute (Johnson and Mathews, 2001; Miguel et al., 2005). In making the comparison, all attributes that differ between the alternatives must have a strict ordering of preferences. That is, more of an attribute must always be either better or worse. For categorical attributes, all respondents must impose the same preference ordering over the attribute levels. Choice of an alternative that is dominated by another alternative is not plausible. A rational response would offer no

[10] Realistic ranges may well have a larger domain than current experiences or market offerings. Indeed, this is one of the advantages of SC data over revealed preference data.

information about the relative importance of each of the attributes, so choice tasks with dominance will contribute little to the estimated model. Consequently, such choice tasks are typically prevented when generating the experimental design.

As described, dominance checks rely on an assumption that all attributes will be attended to. To the best of the author's knowledge, no studies have deviated from this assumption. However, choice tasks exhibiting dominance under full attribute attendance might reduce to a tie across alternatives, once some pattern of ANA is applied. Alternatively, choice tasks that do not exhibit dominance under full attendance may do so conditional on ANA.

Consider the first of these two possibilities, taking the choice task in Table 7.1 as an example. In this choice task, Flight A dominates Flight B, as the two alternatives tie on fare and seat pitch, but Flight A has IFE, which is desirable. Reasonably, a respondent may be indifferent to IFE. In this case, Flight A no longer dominates Flight B, the two flights are equally desirable, and choice between them would be random, ceteris paribus. Admittedly, the IFE attribute might be used as a 'tie breaker', but under true indifference, the alternatives will be indistinguishable. This interaction of ANA and dominance is somewhat irrelevant, as the choice task is of little use either with dominance, or with random choice, and should be be discarded, regardless of ANA.

Table 7.1: Dominance and ANA, example one

	Flight A	**Flight B**	**Ignored?**
Fare	$800	$800	No
Seat pitch	32 inches	32 inches	No
In-flight entertainment	Yes	No	Yes

The second possibility, that dominance can be introduced into a choice task once the respondent decides to only attend to a subset of attributes, is more interesting, and of more concern. Table 7.2 illustrates with an example. Once again there is a tie on price, and while Flight A outperforms Flight B on IFE, we consider the case in which an individual is ignoring IFE. This time, Flight B outperforms Flight A on seat pitch, and this attribute is attended to. Under full attendance, there is no dominance, as the two flights each outperform the other on one attribute. However, once IFE is ignored, Flight B dominates Flight A. Here, dominance is conditional on the ANA behaviour exhibited.

The consequence of this ANA conditional dominance (ANACD) is that the efficiency of the design may decrease as the ANA rates increase. Also, there may be an issue with the scale of the model. The choice probabilities for dominated and dominating alternatives should be

Table 7.2: Dominance and ANA, example two

	Flight A	**Flight B**	**Ignored?**
Price	$800	$800	No
Seat pitch	32 inches	34 inches	No
In-flight entertainment	Yes	No	Yes

zero and one respectively. This would be reflected by the taste coefficients going to infinity. However, the choice tasks with and without dominance are pooled for estimation, and so the coefficients may become biased. More attributes might mitigate ANACD somewhat, as there will be more chance that some attribute breaks the dominance. There may be some way to minimise the incidence of ANACD through appropriate generation of the experimental design, although this may conflict with other measures of design quality, such as the commonly employed *d*-error. Given that SC designs are commonly blocked such that each respondent only sees some of the total set of choice tasks, minimising the incidence of ANACD may be important not just overall, but across the set of choice tasks that each respondent receives. The risk is that a respondent, imposing some pattern of ANA, might have ANACD in all of their choice tasks. It would be better to spread ANACD across the blocks evenly, for each ANA pattern. This special type of dominance remains an area for future work.

This thesis has not investigated the specific interaction of ANA and the properties of individual choice tasks, although a number of studies have. DeShazo and Fermo (2004) estimated the propensity to attend to each attribute (refer to Section 2.2.2, p.51 for details). A number of influences on this propensity were examined, including the cost of evaluation, operationalised by a measure of the standard deviation of the attribute levels. Cameron and DeShazo (2011) also estimated variations in the propensity to attend, as a function of the tradeoffs in the choice tasks. They suggested that the greater the difference in an attribute across alternatives (own-attribute utility difference), the more likely it is to be attended to. Further, there are more benefits to attending to an attribute when the differences in utility between alternatives, excluding that attribute, are not distinct (other-attribute utility difference). Thus, they suggested that the specific tradeoffs in a choice task will influence the propensity to attend, and so ANA might vary across choice tasks. This is consistent with a body of evidence from studies that have compared serial with choice task ANA (refer to Section 2.2.3, p.55 for details).

It seems plausible that a link will exist between preferences and ANA that is induced by some aspect of the choice task. Cameron and DeShazo (2011, p.82) suggested that "the greater the true marginal utility associated with this attribute, the greater the likelihood that it will receive more attention". Notably, the methodology proposed by Cameron and DeShazo (2011) only relies upon an average sensitivity to the attribute when leveraging this insight as a part of their own and other-attribute utility difference measures. If there is preference heterogeneity in the sample, the sensitivities will vary across individuals, and consequently so may the likelihood of that attribute being attended to or not.

An alternative model, suggested only in broad, conceptual terms here, could take this concept further. As with DeShazo and Fermo (2004), Hensher (2006a) and Cameron and DeShazo (2011), the propensity to not attend to an attribute might be a function of some properties of the choice task, such as the number of attributes and alternatives, and the difficulty of evaluating the alternatives. However, attributes that are of low importance *for any given individual* would be prioritised for nonattendance[11]. The model would capture preference heterogeneity, with ANA linked to this preference *heterogeneity*, rather than some average sensitivity. One challenge might be the tension between components of the model that remain invariant across all choice tasks for the individual, and components that vary between each choice task. The taste sensitivities might be an example of the former, while the probability of ANA would be an example of the latter. There may be scope to parameterise the ANA assignment probabilities in the RPANA model as a function of choice task characteristics, as well as the random taste sensitivities, but the specifics of the resulting integration would require careful thought. Behaviourally, the motivation here is to overcome the assumption of consistent ANA behaviour along the panel, whilst still leveraging the assumption that tastes are likely to remain invariant for each individual across choice tasks in an SC experiment.

The RPANA model developed and explored in this thesis treats ANA largely statistically. It is not a formal theoretical model. There is scope to capture some behavioural drivers of ANA, through the nonattendance covariates. Nonetheless, it is not an intrinsically behavioural model, and makes no prescriptions as to what is causing nonattendance. This contrasts with the approach of a limited number of papers in the ANA literature, most notably that of Cameron and DeShazo (2011). They present a formal behavioural theoretical model that at its core considers both the costs and benefits of attention.

The use of a conceptual framework to provide a strong theoretical underpinning to a model that handles ANA is appealing in a number of ways. It may provide more insight

[11] I thank Caspar Chorus for his comments at the 2012 IATBR conference in Toronto, which fuelled this line of thinking.

into why ANA is occurring. It may also allow for the identification of more of a distinction between nonattendance that is independent of context, and that which may vary across choice occasions for any number of reasons, including the properties of the choice task, the attribute levels, learning and fatigue effects, and the context in which the choice is being made. That is, a more theoretical model may assist in the separation of inattention to an attribute, and the lack of preference for an attribute.

To some extent, some of the benefits of a theoretical model can be achieved with the RPANA model, by using covariates to systematically vary the propensity to not attend. This would allow for a more meaningful interpretation of ANA, and what is driving it. It will also generate a more nuanced suite of WTP measures. As discussed already, however, as currently formulated, the RPANA model cannot handle influences on ANA that vary across choice tasks, due to the panel specification of the model. This further motivates research to overcome such a limitation. Ultimately, both the theoretical and statistical approaches to handling ANA have appeal, and future research may see the two approaches move closer together.

Several more points will be made about experimental design in the context of ANA. Campbell and Lorimer (2009) suggested that experimental designs should be generated to allow attribute processing strategies to be identified, although they did not prescribe how this could be achieved. For techniques such at the RPANA model, that leverage repeated choice tasks per respondent, a longer panel length is likely to help. This remains an important area for future research. The experimental design may well impact on ANA at the choice task level, especially where ANA might stem from the tradeoffs faced, as discussed above. Consequently, any study of either the design influences of ANA or choice task ANA should probably consider both. Finally, even where an experimental design may be influencing ANA, and where this may be problematic, the solution certainly is not to ignore ANA. It is better to handle ANA through some means, and either correct for the influence of the design as much as is possible, or use the results of a pilot study to refine the design of a subsequent stage.

Differences in scale are not handled by the RPANA model. Recall that scale has been considered in several ANA papers (see Section 2.2.2). Given that scale can be interpreted as representing the ability of an individual to choose (de Palma et al., 1994), nonattendance may interact with scale in a complex way. Nonetheless, no definitive findings have been found in the literature. What may be required is a more behaviourally based model that can handle behavioural links between ANA and differences in scale. Scale and ANA remains an interesting area for future research.

Beyond ANA[12], the handling of multiple heuristics or decision rules in econometric choice models has received much attention in recent years (Swait and Adamowicz, 2001; Hensher and Greene, 2010; Hensher and Collins, 2011; Leong and Hensher, 2011; Hess et al., 2012). A natural extension to the ANA literature then is to accommodate further decision rules within the model (e.g. Hensher and Greene, 2010; Hensher et al., 2012b). However, the challenges faced with just two decision rules (expounded below) only grow as more decision rules are added to the model. In the face of such challenges, there must be some sufficient gain to make the endeavour worthwhile. Handling more decision rules may, potentially, better explain the choices, and provide more behavioural insight. Several studies, including this thesis, have shown that preference heterogeneity in RPL models may in fact be capturing various forms of process heterogeneity (Hess et al., 2012; Rose et al., 2012b). Hess et al. (2012) noted that for individuals who enact alternative decision rules, welfare measure may not be readily derived. Whilst seemingly a detractor for handling such decision rules, questions must consequently be asked about the meaningfulness of the welfare measures derived for such individuals in conventional random utility maximisation (RUM) models.

Offsetting such potential gains are a number of risks and challenges. One is an uncertainty as to whether the model's estimate of the decision rule is an accurate representation of that decision rule, or whether it is to some extent approximating some other decision rule. Chapter 4 demonstrated how a RPL model, handling RUM only, can approximate ANA in a detrimental way. Even when ANA was handled in the RPANA model, bounded distributions had to be employed to prevent an identification problem stemming from the competing support of the continuous and discrete distributions. In this case, the problems with unbounded distributions could be pinpointed through difficulties with model estimation. More problematic are situations whereby it is not obvious that such a conflict exists.

The addition of random parameters to the RUM component of the model has caused problems in other studies as well. Hess et al. (2012) found that the share of the alternative heuristics decreased once random preference heterogeneity was accommodated in the RUM class, although the heterogeneity was also decreased from the model that did not handle the alternative heuristics. This suggests that unless appropriately specified, one heuristic might approximate another. Cameron and DeShazo (2011) found that the systematic sources of propensity to attend that were significant when preferences were assumed to be homogenous

[12]Note that we could interpret ANA as a heuristic, especially if it is context dependent, or as an expression of preference heterogeneity, especially if it is applied consistently by an individual.

were no longer so once preference heterogeneity was introduced into their model. Thus, their model of process heterogeneity was interchangeable with a conventional RPL model of taste heterogeneity. It is hard in such circumstances to determine which model is closer to the truth. Cameron and DeShazo (2011) argued that their approach provides more insight than taste heterogeneity that is simply random.

Attribute nonattendance can serve as a building block for a number of other 'more complex' decision rules and heuristics. For example, lexicographic behaviour is the nonattendance to all attributes bar one. It may be more appropriate to handle the more complex heuristics directly. Alternatively, with sufficient flexibility, the handling of ANA only might serve as an adequate approximation of a number of other heuristics. Separating out multiple decision rules in the one econometric model is not a trivial exercise, and is fraught with risks. It is a worthwhile endeavour; however, caution is warranted.

Adding more decision rules will require richer data, and may be parametrically more expensive. The experimental design may play a pivotal role, either preventing or inducing the enaction of the alternative decision rules (Hess et al., 2012). There is an issue of how readily the various heuristics can coexist in the one econometric model. The dominant approach has been through various specifications of the LC model. Indeed, this could be applied to the RPANA model, by adding another layer of latent classes. Alternatively, a heuristic such as the aggregation of common-metric attributes (Layton and Hensher, 2010) could be accommodated directly within the model, by specifying one of the (potentially numerous) ANA assignment models to include combinations of attendance over both the disaggregated and aggregated attributes. Adding more decision rules to the RPANA model will remain an area for future research.

Performing the research contained in this thesis has been an instructive process for the author. A crucial skill that has been gained is the ability to conceptualise and build a model from the ground up, and a recognition of the power of being able to do so. The problems faced have also given the author a greater awareness of the various types of identification problems that can arise, and experience with how they may be detected during model estimation. The usefulness of simulation has become apparent, both for detecting problems and testing solutions. Finally, the comments of one of the examiners have provided a greater appreciation for the differences between various types of models, and in particular between econometric models and theoretical models that have richer behavioural underpinnings.

7.3 Conclusion

This research has made three broad contributions. First, it has deepened the understanding of the the impact of ANA on current choice modelling methodologies. Second, some of the existing methodologies for handling ANA have been critically evaluated. Third, a new methodology has been developed that is more flexible, allows more insights to be gained, and outperforms existing methodologies.

An extensive review of the related literature revealed some recognised problems with and limitations to existing techniques for handling ANA. The testing of the existing methodologies herein revealed further problems and limitations. However, promising frameworks within which to approach the problem were identified. Further, aspirational levels were established in terms of the properties of a proposed new methodology, and the behavioural insights that such a methodology should be able to make.

In evaluating the impact of ANA on RPL models, simulated data was employed, to allow for comparisons between the estimated and true values, over many data points. The findings were reached through inspection, and with SURE and Tobit models. It was found that ANA leads to downward bias in the mean of the taste coefficients, upward bias in the extent of taste heterogeneity, especially when the true value is low, and an increase in implausibly signed coefficients, especially when there is greater true taste heterogeneity. Further evidence has been added to a growing body, that what is frequently identified as taste heterogeneity might in part just be ANA. This will have adverse effects as the model outputs, such as WTP, are applied in any given area. What has been newly established is that the extent of true taste heterogeneity will have a strong influence on parameter bias and sign violation. The research provides a further warning about the dangers of not handling ANA, and provides an additional impetus to handle ANA in research and practice.

The evaluation of the ANA model and censored normal distributions, as methodologies for handling ANA, also relied on simulated data. It was found that the censored normal tends to recover bias in all three aspects of the distribution: mean tastes, taste heterogeneity, and the ANA rate. In the presence of taste heterogeneity, the ANA model overestimates both the mean sensitivity, and the ANA rate. The research has shown that the censored normal may work well, but is restrictive. The ANA model is not recommended, as it is accurate only in very limited circumstances, and severe bias may be introduced. Utilising the model outputs from the ANA model may have deleterious consequences. However, the model does serve as the basis of the RPANA model, which shows much more promise.

The third area in which a contribution has been made is the development of the RPANA model, which combines discrete and continuous random parameters. The contribution is two pronged: a model has been developed that performs well and adds new capabilities; and a deep understanding of this model has been gained. When tested on simulated data, the RPANA model inferred both ANA and taste heterogeneity with a fairly high level of accuracy. In an empirical setting, the model was an improvement on the ANA model, and the RPL model with a variety of distributions, including the censored normal. The RPANA model can handle covariates of ANA, and is not reliant on stated ANA, but can leverage it. Also, the model can be specified in a multitude of ways, to maximise parsimony, without violating independence assumptions, where these assumptions are likely to be context dependent. The deeper understanding of the model has resulted in a set of guidelines for others who may wish to employ the model. The thesis has detailed the necessary conditions to ensure identification, and warned of the potential for further problems. It has been found that correctly specifying the model to handle correlation in ANA is important. Tests for the independence were introduced. Also important is the choice of RP distribution. An incorrect choice may lead to a failure to recover a true point mass at zero. Further, high sensitivities may still impact on low sensitivities and influence the recovered ANA rate, and some distributions may handle this better than others.

The findings must be qualified in a number of ways. The RPANA model is reliant on multiple choice observations per individual, which may be difficult to collect in some circumstances. Serial ANA is more likely to be captured by the model, although this may be an advantage. The model is expensive computationally, and potentially parametrically also, depending on the extent of correlation in ANA. The restrictions on the distributions that can be used are somewhat limiting. Further, it is unlikely that ANA can be completely separated from taste heterogeneity, as to some extent ANA and low sensitivities may still approximate each other, albeit far less so than with the ANA model.

Further research might focus on various extensions to the RPANA model, including the estimation of both ANA and correlated random parameters, and the derivation of more model outputs. Testing the model on more datasets will further instill confidence in the model, and may allow additional insights to be gleaned.

Appendix A

Glossary

ANA

Attribute nonattendance.

ANA model

Attribute nonattendance model.

ANACD

Attribute nonattendance conditional dominance.

ASC

Alternative specific constant.

CANA model

Correlated attribute nonattendance model.

DM model

Discrete mixture model.

EAA model

Endogenous attribute attendance model.

EM algorithm

Expectation-Maximisation algorithm.

EMU

Expected maximum utility.

ECLC model

Equality constrained latent class model.

Hybrid ANA model

An attribute nonattendance model in which ANA is assumed to be independent between some, but not all attributes. Specifically, $1 < A < K^{\star}$.

Hybrid RPANA model

A random parameters attribute nonattendance model in which ANA is assumed to be independent between some, but not all attributes. Specifically, $1 < A < K^{\star}$.

IANA model

Independent attribute nonattendance model.

IFE

In-flight entertainment. Typically a personal screen, potentially with video on demand.

LC approach

Use of the latent class model to handle ANA, by constraining some coefficients to zero. An alternative name for the ANA model.

LC model

Latent class model.

MNL model

Multinomial logit model.

RPANA model

Random parameters attribute nonattendance model.

RPCANA model

Random parameters correlated attribute nonattendance model.

RPIANA model

Random parameters independent attribute nonattendance model.

RP distribution

Random parameter distribution.

RPL model

Random parameters logit model.

RUM

Random utility maximisation.

SC

Stated choice.

SURE model

Seemingly unrelated regression equation model.

VTTS

Value of travel time savings.

WTP

Willingness to pay.

Appendix B

Notation

a	A specific ANA assignment model.
A	The number of ANA assignment models.
b	The truncation point of the truncated normal distribution.
c_a	A specific class in the ANA assignment model a.
C_a	The set of realised classes for ANA assignment model a.
h	A specific moment or measure used to describe a distribution.
i	An alternative.
$\vec{i}$	A sequence of choices of alternatives over T choice occasions, $\{i_1, \ldots, i_T\}$.
J	The number of alternatives in a choice task.
k	An attribute.
K	The number of attributes describing the choice alternatives.
$K^\star$	The number of attributes describing the choice alternatives for which attribute nonattendance is modelled.
$K_a^\star$	The number of attributes for which ANA is controlled by ANA assignment model a.
l	An attribute level.
L	The number of levels that an attribute possesses.
$\mathrm{LL}(\beta)$	The log-likelihood conditional on coefficients β.
m	A specific latent class. In the RPANA model, this is a class in the final ANA assignment model.

M	The set of all classes. In the RPANA model, this is the set of realised classes in the final ANA assignment model.
n	A respondent or individual.
N	The total number of respondents or individuals.
P_{nc_a}	The probability of respondent n being assigned to class c_a in the ANA assignment model a.
$P_{m\vert n\vec{i}}$	The posterior class assignment probabilities for class m, conditional on the sequence of choices $\vec{i}$ by respondent n.
$P_{n\{c_1,\ldots,c_A\}}$	The probability of respondent n being assigned to each of the ANA assignment classes $\{c_1,\ldots,c_A\}$. An alternative notation for P_{nm}, in recognition that each class m in the final ANA assignment model represents a unique combination of $c_a, \forall a$.
$P_{n\vec{i}}$	The unconditional probability of a sequence of choices, $\vec{i}$, for respondent n.
$P_{nit\vert c_1,\ldots,c_A}$	The probability of alternative i being chosen on choice occasion t, conditional on respondent n being assigned to each of the ANA assignment classes $\{c_1,\ldots,c_A\}$. An alternative notation for $P_{ni\vert m}$, in recognition that each class m in the final ANA assignment model represents a unique combination of $c_a, \forall a$.
$P_{n\vec{i}\vert c_1,\ldots,c_A}$	The probability of a sequence of choices of alternatives by respondent n, conditional on assignment to each of the ANA assignment classes $\{c_1,\ldots,c_A\}$. An alternative notation for $P_{n\vec{i}\vert m}$, in recognition that each class m in the final ANA assignment model represents a unique combination of $c_a, \forall a$.
$P_{n\vec{i}\vert m}$	The probability of a sequence of choices of alternatives by respondent n, conditional on assignment to class m.
P_{nit}	The MNL probability that alternative i will be chosen by respondent n on choice occasion t.
$P_{nit\vert m}$	The probability of alternative i being chosen on choice occasion t, conditional on respondent n being assigned to class m.
P_{nm}	The probability of respondent n being assigned to class m.
r	A specific treatment in the simulations.
s	A specific dataset in the simulations.
S	The total number of datasets used with the simulations.

t A choice occasion.

T The total number of choice occasions.

U The total utility, comprised of the observed and unobserved components.

V The representative, observed component of utility.

x A vector of observed variables.

x_{nit0} A vector of observed variables, for alternative j, on choice occasion t, and respondent n. The variables represent attributes for which ANA is not modelled.

x_{nita} A vector of observed variables, for alternative j, on choice occasion t, and respondent n. The variables represent attributes for which ANA is controlled by ANA assignment model a.

y A vector of binary choice outcomes for every alternative in every choice task.

y_{njt} The binary choice outcome for alternative j by respondent n on choice occasion t.

z_n A vector of the variables that influences assignment of respondent n to class m. Associated with the vector of parameters θ_{nm}.

β A vector of taste coefficients.

β_0 A vector of taste coefficients for the $K - K^{\star}$ attributes for which ANA is not modelled.

β_a A vector of taste coefficients associated with attributes for which ANA is controlled by ANA assignment model a.

β_{c_a} A vector of taste coefficients associated with a specific class of a specific ANA assignment model. The coefficients are drawn from β_a, but may be censored to zero, depending on what pattern of ANA class c_a represents.

β_m A vector of taste coefficients associated with class m.

$\beta_{n\vec{i}}$ The individual-specific parameter estimates, for respondent n, conditioned on their sequence of choices $\vec{i}$.

ϵ_{nit} The unobserved component of utility.

γ_{c_a} A parameter that serves as a constant term, capturing the the assignment to a class c_a in the ANA assignment model a that cannot be explained by other factors.

γ_m	A parameter that serves as a constant term, capturing the the assignment to a class m that cannot be explained by other factors.
λ_{khr}	The true moment or measure h for attribute k in treatment r in the simulations.
λ_{khrs}	The estimated moment or measure h for attribute k in treatment r and dataset s in the simulations.
μ	The mean of a distribution. For the lognormal, censored normal and truncated normal distributions, this is the mean of the underlying normal distribution.
σ	The standard deviation of the normal distribution, or the spread of the uniform and triangular distributions, or the standard deviation of the underlying normal distribution for the lognormal, censored normal and truncated normal distributions.
θ_{nc_a}	A vector of parameters that captures socio-demographic and other influences on the assignment of respondent n to class c_a in the ANA assignment model a.
θ_{nm}	A vector of parameters that captures socio-demographic and other influences on the assignment of respondent n to class m.

References

Alemu, M. H., Morkbak, M. R., Olsen, S. B., Jensen, C. L., 2011. Attending to the reasons for attribute non-attendance in Choice Experiments. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E: Logistics and Transportation Review 39 (3), 229–244.

Arentze, T. A., Dellaert, B. G. C., Chorus, C. G., 2011. Activated needs, mental representation and choice behavior: an integrated modeling approach. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Others, 2010. A view of cloud computing. Communications of the ACM 53 (4), 50–58.

Bajari, P., Fox, J. T., Ryan, S. P., 2007. Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. The American economic review 97 (2), 459–463.

Balcombe, K., Burton, M., Rigby, D., 2011. Skew and attribute non-attendance within the Bayesian mixed logit model. Journal of Environmental Economics and Management 62 (3), 446–461.

Batley, R., Daly, A., 2006. On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. Journal of Mathematical Psychology 50 (5), 456–467.

Ben-Akiva, M., 1973. The structure of travel demand models. Ph.D. thesis, Massachusetts Institute of Technology.

Ben-Akiva, M., Bolduc, D., Bradley, M., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record (1413), 88–97.

Ben-Akiva, M., Swait, J., 1986. The Akaike likelihood ratio index. Transportation Science 20 (2), 133–136.

Bhat, C. R., 1998. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. Transportation Research Part A: Policy and Practice 32 (7), 495–507.

Bliemer, M. C. J., Rose, J. M., 2011. Experimental design influences on stated choice outputs: An empirical study in air travel choice. Transportation Research Part A: Policy and Practice 45 (1), 63–79.

Bradley, M., 1998. Behavioural models of airport choice and air route choice. In: Ortúzar, J. D., Hensher, D., Jara-Diaz, S. (Eds.), Travel behaviour research: updating the state of play. Elsevier, Oxford, pp. 141–159.

Bradley, M., Daly, A., 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21 (2), 167–184.

Brownstone, D., Train, K., 1998. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 89 (1), 109–129.

Bujosa, A., Riera, A., Hicks, R. L., 2010. Combining Discrete and Continuous Representations of Preference Heterogeneity: A Latent Class Approach. Environmental and Resource Economics 47 (4), 477–493.

Cameron, T. A., DeShazo, J. R., 2011. Differential attention to attributes in utility-theoretic choice models. Journal of Choice Modelling 3 (3), 73–115.

Campbell, D., Doherty, E., Hynes, S., van Rensburg, T., 2010a. Combining discrete and continuous mixing approaches to accommodate heterogeneity in price sensitivities in environmental choice analysis. In: 84th Annual Conference of the Agricultural Economics Society, Edinburgh, Scotland.

Campbell, D., Hensher, D. A., Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. Journal of Environmental Planning and Management 54 (8), 1061–1076.

Campbell, D., Hensher, D. A., Scarpa, R., 2012. Cost thresholds, cut-offs and sensitivities in stated choice analysis: identification and implications. Resource and Energy Economics 34 (3), 396–411.

Campbell, D., Hutchinson, W., Scarpa, R., 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. Environmental and Resource Economics 41 (3), 401–417.

Campbell, D., Lorimer, V., Aravena, C., Hutchinson, G., 2010b. Attribute processing in environmental choice analysis: implications for willingness to pay. In: Agricultural Economics Society Annual Conference. Edinburgh.

Campbell, D., Lorimer, V. S., 2009. Accommodating attribute processing strategies in stated choice analysis: do respondents do what they say they do? In: European association of environmental and resource economists annual conference, Amsterdam.

Carlsson, F., Kataria, M., Lampi, E., 2010. Dealing with ignored attributes in choice experiments on valuation of Sweden. Environmental and Resource Economics 47 (1), 65–89.

Caussade, S., Ortúzar, J. D., Rizzi, L. I., Hensher, D. A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B 39 (7), 621–640.

Chorus, C. G., Arentze, T. A., Timmermans, H. J. P., 2008. A Random Regret-Minimization model of travel choice. Transportation Research Part B 42, 1–18.

Cirillo, C., Axhausen, K. W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research Part A: Policy and Practice 40 (5), 444–457.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K., Mukherjee, A., 2003. Modeling aggregate air-travel itinerary shares: logit model development at a major US airline. Journal of Air Transport Management 9, 361–369.

Collins, A. T., Rose, J. M., Hess, S., 2012. Interactive stated choice surveys: a study of air travel behaviour. Transportation 39, 55–79.

Daly, A., Zachary, S., 1978. Improved multiple choice models. In: Hensher, D., Dalvi, M. (Eds.), Determinants of Travel Choice. Saxon House, Sussex.

de Palma, A., Myers, G. M., Papageorgiou, Y. Y., 1994. Rational Choice Under an Imperfect Ability To Choose. The American Economic Review 84 (3), 419–440.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Deshazo, J. R., Fermo, G., 2002. Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management 44 (1), 123–143.

DeShazo, J. R., Fermo, G., 2004. Implications of Rationally-Adaptive Pre-choice Behavior for the Design and Estimation of Choice Models.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., Doherty, M. L., 1989. Process tracing methods: Contributions, problems, and neglected research questions. Organizational Behavior and Human Decision Processes 43 (1), 75–117.

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

Fosgerau, M., Bierlaire, M., 2007. A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B: Methodological 41 (7), 784–794.

Fosgerau, M., Hess, S., 2008. Competing methods for representing random taste heterogeneity in discrete choice models.

Gensch, D. H., 1987. A Two-Stage Disaggregate Attribute Choice Model. Marketing Science 6 (3), 223–239.

George, E., McCulloch, R., 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association 88, 881–889.

Gilbride, T. J., Allenby, G. M., Brazell, J. D., 2006. Models for heterogeneous variable selection. Journal of Marketing Research 43 (3), 420–430.

Gopinath, D. A., 1995. Modeling heterogeneity in discrete choice processes: Application to travel demand.

Greene, W., 1997. Econometric Analysis, 3rd Edition. Prentice Hall, New Jersey.

Greene, W., 2002. LIMDEP version 8.0. Econometric modeling guide, vol. 2. Econometric Software.

Greene, W. H., 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.

Hauser, J. R., Wernerfelt, B., 1990. An Evaluation Cost Model of Consideration Sets. Journal of Consumer Research 16 (4), 393–408.

Hensher, D., 2010. Hypothetical bias, choice experiments and willingness to pay. Transportation Research Part B: Methodological 44 (6), 735–752.

Hensher, D., Collins, A., 2011. Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? Journal of Choice Modelling 4 (1), 62–89.

Hensher, D., Greene, W., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.

Hensher, D., Rose, J., 2009. Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. Transportation Research Part E: Logistics and Transportation Review 45 (4), 583–590.

Hensher, D., Rose, J., Greene, W., 2012a. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. Transportation 39, 235–245.

Hensher, D. A., 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. Journal of Transport Economics and Policy 38 (3), 425–446.

Hensher, D. A., Sep. 2006a. How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics 21 (6), 861–878.

Hensher, D. A., 2006b. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. Environmental and Resource Economics 34 (1), 7–44.

Hensher, D. A., 2006c. The signs of the times: Imposing a globally signed condition on willingness to pay distributions. Transportation 33 (3), 205–222.

Hensher, D. A., 2007. Reducing sign violation for VTTS distributions through endogenous recognition of an individual's attribute processing strategy. International Journal of Transport Economics 34 (3), 333–349.

Hensher, D. A., 2008. Joint Estimation of Process and Outcome in Choice Experiments and Implications for Willingness to Pay. Journal of Transport Economics and Policy 42, 297–322.

Hensher, D. A., Collins, A. T., Greene, W. H., 2012b. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. Transportation.

Hensher, D. A., Greene, W. H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empirical Economics 39 (2), 413–426.

Hensher, D. A., Layton, D., 2010. Parameter transfer of common-metric attributes in choice analysis: implications for willingness to pay. Transportation 37 (3), 473–490.

Hensher, D. A., Louviere, J. J., 1983. Identifying individual preferences for international air fares: an application of functional measurement theory. Journal of Transport Economics and Policy 17 (3), 225–245.

Hensher, D. A., Rose, J., Bertoia, T., 2007. The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies. Transportation Research Part E 43, 73–89.

Hensher, D. A., Rose, J., Greene, W. H., 2005. The implications on willingness to pay of respondents ignoring specific attributes. Transportation 32, 203–222.

Hensher, D. A., Stopher, P. R., Louviere, J. J., 2001. An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application. Journal of Air Transport Management 7 (6), 373–379.

Hess, S., 2007. Posterior analysis of random taste coefficients in air travel behaviour modelling. Journal of Air Transport Management 13 (4), 203–212.

Hess, S., Adler, T., Polak, J. W., 2007a. Modelling airport and airline choice behaviour with the use of stated preference survey data. Transportation Research Part E: Logistics and Transportation Review 43 (3), 221–233.

Hess, S., Bierlaire, M., Polak, J. W., 2007b. A systematic comparison of continuous and discrete mixture models. European Transport 37, 35–61.

Hess, S., Hensher, D. A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B: Methodological 44 (6), 781–790.

Hess, S., Rose, J., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B: Methodological 43 (6), 708–719.

Hess, S., Rose, J. M., 2007. A latent class approach to modelling heterogeneous information processing strategies in SP studies. In: Oslo Workshop on Valuation Methods in Transport Planning.

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 (5), 847–863.

Hess, S., Rose, J. M., Polak, J., 2010. Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transportation Research Part D: Transport and Environment 15 (7), 405–417.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., Caussade, S., 2011. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity.

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.

Hole, A. R., 2011a. A discrete choice model with endogenous attribute attendance. Economics Letters 110 (3), 203–205.

Hole, A. R., 2011b. Attribute non-attendance in patients' choice of general practitioner appointment. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Hole, A. R., Kolstad, J. R., Gyrd-Hansen, D., 2012. Inferred vs stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour.

Hoyos, D., Mariel, P., Meyerhoff, J., 2010. Comparing the performance of different approaches to deal with attribute non-attendance in discrete choice experiments: a simulation experiment.

Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice designs. Journal of Marketing research, 307–317.

Johnson, F. R., Mathews, K. E., 2001. Sources and effects of utility-theoretic inconsistency in stated-preference surveys. American Journal of Agricultural Economics 83 (5), 1328–1333.

Johnson, N. L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36 (1/2), 149–176.

Johnson, R., 2000. Monotonicity constraints in choice-based conjoint with hierarchical bayes.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society 47 (2), 263–291.

Kamakura, W. A., Russell, G. J., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26 (4), 379–390.

Kanafani, A., Sadoulet, E., 1977. The partitioning of long haul air traffic–A study in multinomial choice. Transportation Research 11 (1), 1–8.

Kaye-Blake, W., Abell, W., Zellman, E., 2009. Respondents' ignoring of attribute information in a choice modelling survey. Australian Journal of Agricultural and Resource Economics 53 (4), 547–564.

Kjaer, T., Bech, M., Gyrd-Hansen, D., Hart-Hansen, K., 2006. Ordering effect and price sensitivity in discrete choice experiments: need we worry? Health economics 15 (11), 1217–1228.

Lancaster, K., 1966. A new approach to consumer theory. The Journal of Political Economy 74 (2), 132–157.

Layton, D., Hensher, D. A., 2010. Aggregation of common-metric attributes in preference revelation and implications for willingness to pay. Transportation Research Part D: Transport and Environment 15 (7), 394–404.

Lenk, P., DeSarbo, W., 2000. Bayesian inference for finite mixtures of generalized linear models with random effects. Psychometrika 65 (1), 93–119.

Leong, W., Hensher, D. A., 2011. Embedding Multiple Heuristics into Choice Models: An Exploratory Analysis. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Leong, W., Hensher, D. A., 2012. Embedding Decision Heuristics in Discrete Choice Models: A Review. Transport Reviews 32 (3), 313–331.

Louviere, J., Street, D., Burgess, L., Wasi, N., Islam, T., Marley, A., 2008. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. Journal of Choice Modelling 1 (1), 128–163.

Mariel, P., Meyerhoff, J., Hoyos, D., 2011. Stated or inferred attribute non-attendance? A simulation approach. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Marschak, J., 1960. Binary-choice constraints and random utility indicators. In: Arrow, K. (Ed.), Stanford Symposium on Mathematical Methods in the Social Sciences. Stanford University Press, Stanford, pp. 312–329.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

McFadden, D., 1978. Modeling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam, pp. 75–96.

Meyerhoff, J., Liebe, U., 2009. Discontinuous preferences in choice experiments: Evidence at the choice task level. In: 17th European Association of Environmental and Resource Economists Conference. Amsterdam.

Miguel, F. S., Ryan, M., Amaya-Amaya, M., 2005. 'Irrational' stated preferences: a quantitative and qualitative investigation. Health economics 14 (3), 307–322.

Mokhtarian, P. L., Salomon, I., 2001. How derived is the demand for travel? Some conceptual and measurement considerations. Transportation Research Part A: Policy and Practice 35 (8), 695–719.

Nisbett, R. E., Wilson, T. D., 1977. Telling more than we can know: Verbal reports on mental processes. Psychological Review 84 (3), 231.

Ortúzar, J. D., Simonetti, C., 2008. Modelling the demand for medium distance air travel with the mixed data estimation method. Journal of Air Transport Management 14 (6), 297–303.

Payne, J. W., Bettman, J. R., Coupey, E., Johnson, E. J., 1992. A constructive process view of decision making: Multiple strategies in judgement and choice. Acta Psychologica 80 (1-3), 107–141.

Payne, J. W., Bettman, J. R., Johnson, E. J., 1988. Adaptive strategy selection in decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition 14 (3), 534–552.

Proussaloglou, K., Koppelman, F., 1995. Air carrier demand. Transportation 22 (4), 371–388.

Proussaloglou, K., Koppelman, F. S., 1999. The choice of air carrier, flight, and fare class. Journal of Air Transport Management 5 (4), 193–201.

Puckett, S. M., Hensher, D. A., 2008. The role of attribute processing strategies in estimating the preferences of road freight stakeholders. Transportation Research Part E 44, 379–395.

Puckett, S. M., Hensher, D. A., 2009. Revealing the extent of process heterogeneity in choice analysis: An empirical assessment. Transportation Research Part A: Policy and Practice 43 (2), 117–126.

Revelt, D. A., 1999. Three discrete choice random coefficients papers and one police-crime study. Ph.D. thesis, University of California, Berkeley.

Roberts, J. H., Lattin, J. M., 1991. Development and Testing of a Model of Consideration Set Composition. Journal of Marketing Research 28, 429–440.

Roberts, J. H., Lattin, J. M., 1997. Consideration: Review of Research and Prospects for Future Insights. Journal of Marketing Research 34, 406–410.

Rose, J. M., 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.

Rose, J. M., Bliemer, M., Hensher, D. A., Collins, A. T., 2008. Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B: Methodological 42 (4), 395–406.

Rose, J. M., Hensher, D. A., Caussade, S., Ortúzar, J. D., Jou, R. C., 2009. Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. Journal of Transport Geography 17 (1), 21–29.

Rose, J. M., Hensher, D. A., Greene, W. H., 2005. Recovering costs through price and service differentiation: Accounting for exogenous information on attribute processing strategies in airline choice. Journal of Air Transport Management 11, 400–407.

Rose, J. M., Hensher, D. A., Greene, W. H., Washington, S. P., 2012a. Attribute exclusion strategies in airline choice: accounting for exogenous information on decision maker processing strategies in models of discrete choice. Transportmetrica 8 (5), 344–360.

Rose, J. M., Hess, S., Collins, A. T., 2012b. What if my model assumptions are wrong: the impact of non-standard behaviour on choice model estimation. Journal of Transport Economics and Policy.

Russo, J. E., Dosher, B. A., 1983. Strategies for multiattribute binary choice. Journal of Experimental Psychology: Learning, Memory, and Cognition 9 (4), 676–696.

Saelensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological Economics 59 (3), 331–340.

Scarpa, R., Drucker, A. G., Anderson, S., Ferraes-Ehuan, N., Gomez, V., Risopatron, C. R., Rubio-Leonel, O., 2003. Valuing genetic resources in peasant economies: the case of 'hairless' creole pigs in Yucatan. Ecological Economics 45 (3), 427–443.

Scarpa, R., Gilbride, T. J., Campbell, D., Hensher, D. A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. European Review of Agricultural Economics 36 (2), 151–174.

Scarpa, R., Raffaelli, R., Notaro, S., Louviere, J., 2011. Modelling The Effects Of Stated Attribute Non-Attendance On Its Inference: An Application to Visitors Benefits from the Alpine Grazing Commons. In: 2nd International Choice Modelling Conference, Leeds, United Kingdom.

Scarpa, R., Thiene, M., Hensher, D. A., 2010. Monitoring Choice Task Attribute Attendance in Nonmarket Valuation of Multiple Park Management Services: Does It Matter? Land Economics 86 (4), 817–839.

Shocker, A. D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions. Marketing Letters 2 (3), 181–197.

Simon, H. A., 1956. Rational choice and the structure of the environment. Psychological Review; Psychological Review 63 (2), 129–138.

Simon, H. A., 1972. Theories of bounded rationality. Decision and organization 1, 161–176.

Sims, C. A., 2003. Implications of rational inattention. Journal of Monetary Economics 50 (3), 665–690.

Sims, C. A., 2010. Rational inattention and monetary economics. Handbook of Monetary Economics 3.

Stewart, N., Chater, N., Stott, H. P., Reimers, S., 2003. Prospect relativity: how choice options influence decision under risk. Journal of Experimental Psychology: General 132 (1), 23.

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

Swait, J., Adamowicz, W., 2001. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. Journal of Consumer Research 28 (1), 135–148.

Swait, J., Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. Transportation Research Part B 21 (2), 91–102.

Theis, G., Adler, T., Clarke, J. P., Ben-Akiva, M., 2006. Risk aversion to short connections in airline itinerary choice. Transportation Research Record: Journal of the Transportation Research Board 1951, 28–36.

Train, K., 2001. A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. University of California, Berkeley.

Train, K., 2009. Discrete choice methods with simulation, 2nd Edition. Cambridge University Press, New York.

Train, K., Sonnier, G., 2005. Mixed logit with bounded distributions of correlated partworths. In: Scarpa, R., Alberini, A. (Eds.), Applications of simulation methods in environmental and resource economics. Springer, pp. 117–134.

Train, K. E., 1998. Recreation demand models with taste differences over people. Land economics 74 (2), 230–239.

Train, K. E., 2008. EM algorithms for nonparametric estimation of mixing distributions. Journal of Choice Modelling 1 (1), 40–69.

Tversky, A., 1972. Elimination by aspects: A theory of choice. Psychological Review 79 (4), 281–289.

Walker, J., 2002. Mixed logit (or logit kernel) model: Dispelling misconceptions of identification. Transportation Research Record: Journal of the Transportation Research Board 1805, 86–98.

Warburg, V., Bhat, C., Adler, T., 2006. Modeling demographic and unobserved heterogeneity in air passengers' sensitivity to service attributes in itinerary choice. Transportation Research Record: Journal of the Transportation Research Board 1951, 7–16.

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