WORKING PAPER

ITLS-WP-13-05

A model for the simultaneous inference of attribute nonattendance and taste heterogeneity.

By
Andrew Collins, John Rose and David Hensher

March 2013

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

WORKING PAPER

ITLS-WP-13-05

A model for the simultaneous inference of attribute nonattendance and taste heterogeneity.

By
Andrew Collins, John Rose and
David Hensher

March 2013

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-13-05

TITLE: **A model for the simultaneous inference of attribute nonattendance and taste heterogeneity**

ABSTRACT: An extensive literature has recognised that when travel choices are made, only a subset of the attributes of the choice alternatives may be considered or attended to by each decision maker. This paper introduces the random parameters attribute nonattendance (RPANA) model. Attribute nonattendance (ANA) is estimated through the model, which does not rely on stated ANA, although the latter can be introduced as a covariate, in recognition that whilst concerns have been raised about the reliability of stated ANA, it may provide the analyst with valuable information. The model employs a latent class structure to capture an elevated mass of taste coefficients at zero; a technique widely employed in the literature. However, preference heterogeneity can additionally be captured as a continuous distribution, with random parameters. The latent class component introduced in this model is highly flexible and parsimonious, and can exploit any independence of ANA across the attributes, whilst also handling correlation in ANA across subsets of attributes, as needed. Results are presented from a stated choice experiment investigating short haul flights. Non-zero probabilities of ANA are estimated for all attributes, and the RPANA model represents an improvement on the random parameters logit model in terms of model fit. Specific issues with model identification are discussed, and some potential pitfalls noted.

KEYWORDS: ***Attribute nonattendance, random parameters attribute nonattendance model, latent class model, random parameters logit.***

AUTHORS: **Collins, Rose and Hensher**

CONTACT: INSTITUTE of TRANSPORT and LOGISTICS STUDIES (C13)
The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1824
Facsimile: +612 9114 1722
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

DATE: March 2013

1. Introduction

A stream of literature has developed in recent years that recognises that when travel choices are made, each attribute of each choice alternative may only be considered by a subset of those making the choice. This has been described as the ignoring of attributes (Hensher et al., 2005), and, as used in this paper, attribute nonattendance (ANA; Scarpa et al., 2009).

One interpretation of ANA is that it is a valid phenomenon reflecting the preferences of the individual making a choice. Cirillo and Axhausen (2006) suggest 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) recognise that consumers choosing a product might have no intrinsic value for some of the attributes of the products on offer. This is plausible for travel choices such as the selection of long haul flights, where the choice alternatives may have many features (in-flight entertainment, seat pitch, stopover duration etc), and each individual may only value some of these features (Collins et al., 2012). In contexts such as these, ANA is not inherently a problem, but a valid behavioural phenomenon that may be of interest to the analyst. However, failure to capture ANA may lead to biases in model outputs such as willingness to pay measures (Hensher et al., 2005), and it may contribute to implausibly signed random parameter coefficients (Hensher, 2007). A robust method for identifying ANA is important, and such a method is a key contribution of this paper.

Another interpretation is that, for various reasons, what we believe to be ANA is only a partial reflection of the preferences of an individual. This may be due to the experimental design of a stated choice experiment. Hensher et al. (2012a) note that ANA might be a result of presenting respondents with behaviourally questionable trade-offs. Various studies have questioned the reliability of statements made by respondents about which attribute they ignore (e.g. Hess and Rose, 2007; Carlsson et al., 2010). An alternative is to identify ANA econometrically, however many such applications have resulted in implausibly high ANA rates, and the popular latent class approach (Hess and Rose, 2007) has been recently criticised for confounding ANA and preference heterogeneity (Hess et al., 2012). The ANA literature has paid little attention to the censored normal random parameter distribution (Train and Sonnier, 2005) as a means of capturing ANA. However, since the estimated moments of the distribution capture both ANA and preference heterogeneity, the two phenomena could also be confounded. Hess and Hensher (2010) propose a technique that is informed by the conditional parameters estimates, however it is reliant on the selection of an arbitrary threshold value, and Mariel et al. (2011) have shown that the most accurate such value is dependent on the true ANA rate, which is latent.

This paper presents the random parameters attribute nonattendance (RPANA) model. The model analytically captures ANA, and is an extension of the latent class approach prevalent in the literature, which will be referred to as the attribute nonattendance (ANA) model. As with the ANA model, respondents belong to latent classes up to a probability, where the classes represent certain combinations of attendance and nonattendance to the attributes. Unlike most existing work, the taste coefficients are specified as a distribution, so capturing the preference heterogeneity of those that attend to the attribute. The approach is similar to that proposed by Hess et al. (2012), however it extends upon this work in several ways detailed below[1]. Hensher et al. (2012b) have also combined the latent class model with random parameters, in the context of both ANA and the aggregation of common-metric attributes, but only found a negligible improvement in model fit.

The ANA model has been proposed in two forms. In one, a conventional latent class assignment model determines the probability of a respondent being assigned to a specific combination of ANA

[1]The model has been developed simultaneously, but independently.

across the attributes (Scarpa et al., 2009). An alternative approach proposed by Hole (2011) estimates the probability of nonattendance to each attribute directly, and is more parsimonious, but assumes that ANA is independent across attributes. The RPANA model generalises these two approaches, and can handle varying degrees of correlation in ANA across the attributes. This maximises parsimony, which is econometrically advantageous, but only to the extent that it is supported by the ANA behaviour. For some attributes in the empirical application contained herein, it is found that modelling ANA does not lead to an improvement over the RPL model, unless correlation in ANA is allowed, or covariates are introduced to vary the probability of ANA across respondents. Stated ANA is used as a covariate, providing further insights into the reliability of stated ANA. Limitations in what distributions can be employed by the RPANA model are noted, and it is recommended that effects coding be employed over dummy coding for attributes for which ANA is modelled.

2. Methodology

2.1 Attribute nonattendance (ANA) model

The model presented in this section generalises the latent class approach to modelling ANA (Hess and Rose, 2007; Hole, 2011). The two existing approaches are first broadly outlined, and then a generalised model is introduced in detail. Since the two existing approaches are special cases of the generalised model, the latter will be used to precisely define the former.

Consider a choice task wherein the choice 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 always attended to, or econometrically motivated, to lessen the number of parameters that must be estimated.

Under the latent class 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 imposing 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 latent class model. Define M as the number of classes, where $M = 2^{K^\star}$ if all 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[2]. In some 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

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

calculate these probabilities. The choice alternatives are described by K attributes, $K^\star$ of which we model attendance or nonattendance to. 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 by multiplying the final ANA assignment probabilities by the choice probabilities that are 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 latent class 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.

An alternative, more parsimonious approach for generating the final ANA assignment probabilities has been proposed by Hole (2011). 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[3]. 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 latent class 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 conventional latent class approach.

Whether the attribute nonattendance 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 Akaike Information Criterion (AIC)[4]. 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.

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

The conventional latent class approach makes no such assumption, and can handle any correlation structure over the $K^\star$ attributes. The conventional latent class 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, assuming 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, 2011), 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 $K_a^\star$ attributes. If all combinations of nonattendance to the $K_a^\star$ attributes are to be modelled, ANA assignment model a will have $2^{K_a^\star}$ classes. Specific combinations of attendance can be excluded by the analyst, resulting in fewer classes. Define C_a as the realised number of classes for each a. The final ANA assignment model will have $M = \prod_a^A C_a$ classes.

Define P_{nac} as the probability of respondent n belonging to class c in ANA assignment model a. The probability is calculated with an MNL model, such that

$$P_{nac} = \frac{e^{(\gamma_{ac}+\theta_{nac}z_n)}}{\sum_d^{C_a} e^{(\gamma_{ad}+\theta_{nad}z_n)}}. \tag{1}$$

A parameter, γ_{ac}, serves as a constant term, capturing the assignment to class c that cannot be explained by other factors. A vector of parameters, θ_{nac}, captures socio-demographic and other influences on assignment to class c in ANA assignment model a, for respondent n. To ensure identification, γ_{ac} is constrained to zero for one class. Given that most of the discussion in the literature is around attribute *non*attendance, constraining γ_{ac} 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 of the C_a classes 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 will represent a unique pattern of ANA over all $K^\star$ attributes for which ANA is modelled. This pattern of ANA will be represented by a unique set of ANA assignment model classes, $\{c_1, \ldots, c_A\}$. The probability of respondent n belonging to class m is

$$P_{nm} = P_{n\{c_1,\ldots,c_A\}} = \prod_a^A P_{nac_a}. \tag{2}$$

Substituting in Equation 1, this becomes

$$P_{n\{c_1,\ldots,c_A\}} = \prod_a^A \frac{e^{(\gamma_{ac_a}+\theta_{nac_a}z_n)}}{\sum_d^{C_a} e^{(\gamma_{ad}+\theta_{nad}z_n)}}. \tag{3}$$

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 multinomial logit 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, before being substituted by the RPL model in the next section.

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_{ni}, which is composed of the representative utility V_{ni}, and the unobserved component of utility, ϵ_{ni}. The representative component is associated with a vector of observed variables, x_{ni}. The utility associated with these variables is estimated with a vector of taste coefficients β, such that the representative utility is $V_{ni} = \beta x_{ni}$. For the MNL model, the probability that alternative i will be chosen is

$$P_{ni} = \frac{e^{\beta x_{ni}}}{\sum_j^J e^{\beta x_{nj}}}. \tag{4}$$

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 coefficients 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. Each combination will represent a unique pattern of censoring of β_a. For each a, introduce C_a sets, each denoted β_{ac}. The elements of β_{ac} 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_{nj}, are similarly partitioned into $A + 1$ subsets. Variables associated with attributes for which ANA is not modelled are in set x_{nj0}, while the variables associated with attributes for which ANA is modelled are partitioned into A subsets x_{nja}.

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

$$V_{nj|c_1,\ldots,c_A} = \beta_0 x_{nj0} + \sum\nolimits_a^A \beta_{ac_a} x_{nja}. \tag{5}$$

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, conditional on assignment to classes $\{c_1, \ldots, c_A\}$, is

$$P_{ni|c_1,\ldots,c_A} = \frac{e^{\beta_0 x_{ni0} + \sum_a^A \beta_{ac_a} x_{nia}}}{\sum_j^J e^{\beta_0 x_{nj0} + \sum_a^A \beta_{ac_a} x_{nja}}}. \tag{6}$$

For panel data, we can specify the probability with respect to a sequence of choices of alternatives over T time periods, $\{i_1, \ldots, i_T\}$. 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

$$P_{n|c_1,\ldots,c_A} = \prod_t^T \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{ac_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{ac_a} x_{njta}}} \right]. \tag{7}$$

The unconditional probability of a sequence of choices 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

$$P_n = \sum_{c_1}^{C_1} \cdots \sum_{c_A}^{C_A} P_{n\{c_1,\ldots,c_A\}} P_{n|c_1,\ldots,c_A}. \tag{8}$$

Substituting in Equations 3 and 7, Equation 8 becomes

$$P_n = \sum_{c_1}^{C_1} \cdots \sum_{c_A}^{C_A} \prod_a^A \left[\frac{e^{(\gamma_{ac_a} + \theta_{nac_a} z_n)}}{\sum_d^{C_a} e^{(\gamma_{ad} + \theta_{nad} z_n)}} \right] \prod_t^T \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{ac_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{ac_a} x_{njta}}} \right]. \tag{9}$$

Certain specifications of A allow the model to represent the two latent class approaches in the literature. If there is only one ANA assignment model, i.e. $A = 1$, then this is a conventional latent class 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 endogenous attribute attendance model from Hole (2011). 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 ANA model has not been presented in the literature, and represents one of the contributions of this paper. 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. 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).

2.2 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 paper employs several distributions for which a single moment is estimated; notably, the constrained triangular and uniform distributions, wherein the spread is constrained to equal the mean, and the Rayleigh distribution.

Equation 7 now becomes

$$P_{n|c_1,\ldots,c_A} = \int \prod_t^T \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{ac_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{ac_a} x_{njta}}} \right] f(\beta) d\beta. \tag{10}$$

Substituting Equation 10 into Equation 8, we obtain an unconditional probability of a sequence of choices for respondent n of

$$P_n = \sum_{c_1}^{C_1} \cdots \sum_{c_A}^{C_A} \prod_{a}^{A} \left[\frac{e^{(\gamma_{ac_a}+\theta_{nac_a} z_n)}}{\sum_d^{C_a} e^{(\gamma_{ad}+\theta_{nad} z_n)}} \right] \int \prod_t^T \left[\frac{e^{\beta_0 x_{ni_t t0} + \sum_a^A \beta_{ac_a} x_{ni_t ta}}}{\sum_j^J e^{\beta_0 x_{njt0} + \sum_a^A \beta_{ac_a} x_{njta}}} \right] f(\beta) d\beta. \tag{11}$$

This choice probability underpins the RPANA model. Practical issues with estimating the RPANA model are discussed in Section 4.3.1.

2.3 Summary of ANA models in the literature

Table 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 paper (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 are not set to zero; and whether covariates were introduced to vary the ANA probabilities across respondents.

Table 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 (2011)	IANA	5	No	Yes	-
Hess et al. (2012)	RPIANA	2/5/6	Yes	Yes	-
Hensher et al. (2012b)	RPCANA	3	Yes	No	-
This paper	RPANA generalised	4	Yes	Yes	Stated ANA

3. Empirical setting

The empirical setting for this paper is a stated choice 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. 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 paper utilises the first 20 choice tasks completed by each respondent. The sample consisted of students, with an average age of 21. Fifty nine percent of the sample was female, and 41 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[5], and so these three observations are dropped. Interestingly, this is an extreme case of attribute nonattendance, where all attributes are ignored, and only the airline labels are attended to. The final sample size is 4200 observations across 210 respondents. Six point nine percent of these respondents 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.

4. Results

4.1 MNL model

The first model estimated is an MNL model, which is reported in Table 2. 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 2 and henceforth); and WTP to avoid an undesirable attribute level, or a one unit increase in an undesirable attribute ($^-$). 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. The departure time 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

[5] Panel lengths in stated choice experiments are typically shorter. Bliemer and Rose (2011) examined top tier 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.

paribus.

Table 2: MNL model

	Param.	***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			
Parameters	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.

Alternative specific constants (ASCs) 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 willingness to pay 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, 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. Consequently, when modelling ANA, a censoring of the ASCs to zero will be interpreted as nonattendance to the airline, even though there may be some degree of confounding with the other unobserved effects. In the interest of brevity, any reference in the text to nonattendance to attributes also includes nonattendance to the choice alternative labels.

Model fit statistics appear reasonable, and serve foremost as a baseline for subsequent models. All models estimated in this paper 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.

4.2 ANA model

This section presents the results from CANA and IANA models, with nonattendance modelled for fare, flight time, departure time, and airline. In the CANA model, any pattern of correlation of ANA can be captured, while the IANA model assumes independence of ANA. The plausibility of the ANA rates will be assessed. Model outputs and fit will be compared between the two, to gauge whether the assumption of independence holds. Finally, the models will serve as one benchmark for the RPANA models. As discussed in Section 2., nonattendance to dummy coded attributes (departure time) and ASCs (airline) is best modelled by setting to zero the coefficients for *all* dummy coded levels or ASCs. If not all coefficients are set to zero, then what is captured is not nonattendance, but rather an alternative expression of preference.

Table 3 presents the model results. There is a strong alignment in nonattendance rates between the two ANA models, with the exception of flight time, with a rate of 16.33 percent for the CANA model and 11.78 percent for the IANA model. However, the estimated rates themselves do not always have face validity, which in part can be established by a comparison with the stated ANA rates[6]. The estimated ANA rates of fare, at 13.63 and 13.26, are somewhat higher than the stated ANA rate of 6.9 percent. The ANA rate of flight time, estimated by the CANA model at 16.33 percent, is close to the stated rate of 18.1 percent. However, the rate of 11.78 percent estimated by the IANA model is somewhat lower. The estimated and stated rates are wildly divergent for departure time, with estimated rates of 53.42 and 52.17 being far higher than the stated rate of 15.95 percent. The stated rate of attendance to the airline alternative labels was not collected in the survey, however the estimated rates of 78.98 and 80.72 percent appear implausibly high. In sum, the ANA rates appear questionable.

Table 3: ANA models accommodating ANA for all attributes

Methodology	**CANA**			**IANA**		
	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		
Parameters	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.

[6]While stated ANA is not reliable, it provides a ballpark figure.

Since a single parameter controls the ANA rate for each attribute in the IANA model, the associated standard error can be used to provide a measure of statistical reliability of the ANA rate. Table 3 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.

Comparing model fits, both models offer a significant improvement on the MNL model, with drastically lower AIC values. The CANA model has a better log likelihood value than the IANA model, and outperforms it on the AIC despite costing an additional 11 parameters. One possible cause is a violation of the assumption of independence of ANA. The CANA model makes no such assumption between *any* attributes, but is parametrically expensive. The generalised ANA model introduced in this paper allows the assumption to be made only for subsets of attributes. Given the shortcomings of the ANA model, the generalised approach will be explored in the context of the RPANA model only. Indeed, the introduction of extra parameters as random coefficients are introduced will likely make a parsimonious model specification even more desirable.

4.3 RPANA model

4.3.1 Model specification and identification

For the RPANA models, 5000 Halton draws are employed. 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 random parameter 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. In this empirical context, any attempt to include such distributions led to a multitude of estimation problems, including flat log likelihoods 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. 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, at least in the context of this dataset, 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 do not encounter problems in this dataset include the constrained triangular, lognormal, and Rayleigh.

However, the use of a zero bound distribution appears to be a necessary but not sufficient condition. Problems were encountered in this dataset 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 attribute nonattendance and low attribute sensitivity.

The problem with distributions spanning zero poses a challenge in this empirical setting for the dummy coded departure time parameters and alternative specific constants, all of which span zero when RPL models are estimated, and have behaviourally sound justifications for doing so. For example, most respondents have a preference against 6am departures, but some do not. Consequently, the identification problem presented above might apply here, although it appears to be overcome by the joint censoring of each of the coefficients associated with an attribute, when that attribute is ignored. 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 or alternative label. Dummy coding the departure time and airline, and introducing ANA jointly across all related parameters, is found to provide large improvements 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 with L levels is coded into $L-1$ variables. A utility coefficient, β_l, is estimated for each of these variables. The base level of utility is not zero, as with dummy coding, but $\sum_l^{L-1} -\beta_l$. 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.

4.3.2 Single attribute nonattendance

The first set of RPANA models presented will only model nonattendance to a single attribute. This allows the impact of ANA to each attribute to be examined in isolation. Since ANA is modelled for one attribute only, no assumption need be made about the independence or correlation of ANA. Comparisons will be made between the RPANA1 and RPL models. Since an RPL model is nested within the RPANA model if the same random parameter distributions are specified, likelihood ratio tests can be performed to see if the modelling of ANA leads to a statistically significant improvement in model fit. It is also possible that an RPL model with different distributions outperforms the RPANA model. Such comparisons will be made using the AIC. Table 4 first details the RPL model that serves as the base specification for the RPANA1 models, then the four RPANA1 models. Other RPL and RPANA1 models were estimated but are not reported here, and will be drawn upon as required.

Three RPANA1 fare models were estimated, with lognormal, constrained triangular, and Rayleigh distributions for fare. The best fitting model, which uses the lognormal distribution, is the second model presented in Table 4. The ANA rate is low, at 2.12 percent, and statistically different from zero percent (whereupon it would collapse to an RPL model). Confidence intervals provide another useful way to assess the precision of the estimated ANA rate. A 95 percent confidence interval ranges from 0.32 to 12.67 percent. The stated ANA rate of 6.9 percent is higher than the estimated rate, but comfortably within the confidence interval. The ANA rate estimated with the RPANA1 model is much lower than the ANA rates of 13.63 and 13.26 percent inferred from the CANA and IANA models reported earlier. However, the ANA models were found to estimate suspiciously high ANA rates, and the model has recently been criticised for confounding ANA with preference heterogeneity (Hess et al., 2012).

To assess whether modelling ANA to fare leads to an improved model fit, each of the three RPANA1 fare models estimated are first compared with their RPL counterparts using likelihood ratio tests. An improvement in is found for the constrained triangular distribution, with log likelihood improving from -2327.44 to -2323.23, at the cost of one parameter. 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 ($8.40; \chi^2_{1,.05} = 3.84$). However, the null hypothesis that the two models are equivalent cannot be rejected for the Rayleigh distribution ($1.35; \chi^2_{1,.05} = 3.84$) or the lognormal ($1.14; \chi^2_{1,.05} = 3.84$). Furthermore, using the AIC, even the best fitting RPANA1 fare model, with lognormally distributed fare (1.1056), is outperformed by the RPL model with uniformly distributed fare (1.1044). Use of the constrained uniform distribution for fare in a RPANA1 fare model led to identification problems.

A lack in improvement over the best RPL model might stem from characteristics of the fare distributions in both the best RPL and RPANA models. The uniform distribution in the RPL model has a mean of -0.1586 and a spread of 0.1493, meaning that there is some support near zero, which may approximate ANA[7]. The lack of tapering in the distribution over its support also means that there is a reasonable mass which represents more extreme sensitivities. In addition, the properties of the lognormal distribution in the RPANA model might be conflicting with the ANA mass point. 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 mass point 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 mass point at zero. Indeed, the ANA rate with the lognormal distribution (2.12 percent) is lower than with the constrained triangular (3.47 percent), which suggests that such confounding may be occurring.

The RPANA1 flight time model, with a constrained triangular distribution for flight time[8], is presented in Table 4. The ANA rate of 10.78 percent is significantly different to zero, and is close to the IANA model rate of 11.78 percent, and a little less than the CANA model rate of 16.33 percent. However, again, the RPANA model strongly outperforms the ANA models. The RPANA model rate of 10.78 percent is somewhat lower than the stated ANA rate of 18.1 percent, but again the stated rate lies comfortably within the confidence interval, which spans from 4.00 percent to 25.94 percent. The ANA rates vary notably is the distribution for flight time varies. With the lognormal, the ANA rate is 12.96 percent, while for the Rayleigh it is just 4.82 percent. It is possible that some confounding

[7]This may be why the RPANA1 fare model with a constrained uniform distribution cannot be identified.

[8]The Rayleigh distribution performed very slightly better on the AIC, but performed worse in some later models. Thus the constrained triangular distribution is reported, for consistency.

Table 4: RPL and RPANA1 models

		RPL		RPANA1 fare		RPANA1 fl. time		RPANA1 dep. time		RPANA1 airline	
		Param.	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**	**Param.**	***t*-ratio**
Fare	μ	-2.0250	-30.92	-1.9795	-33.53	-2.0257	-29.19	-2.0162	-25.72	-2.0007	-30.03
(lognormal)	σ	0.8618	13.56	0.7906	10.09	0.8586	13.57	0.8982	12.89	0.8534	13.06
	$\gamma_{Ignored}$	-	-	-3.8333	3.17[1]	-	-	-	-	-	-
	$P_{Ignored}$	-		2.12%		-		-		-	
	$P_{Ignored}$ 95% C.I.	-	-	0.32%	12.67%	-	-	-	-	-	-
Flight time	μ	-0.0717	-15.73	-0.0722	-16.25	-0.0833	-16.30	-0.0758	-15.05	-0.0608	-13.08
(varies)	σ	0.0407	8.19	0.0409	8.07	-	-	0.0407	7.57	-1.3928	-10.30
	Distribution	Normal		Normal		Const. triangular		Normal		Normal	
	$\gamma_{Ignored}$	-	-	-	-	-2.1139	8.82[1]	-	-	-	-
	$P_{Ignored}$	-		-		10.78%		-		-	
	$P_{Ignored}$ 95% C.I.	-	-	-	-	4.00%	25.94%	-	-	-	-
Depart 6am	μ	-1.3466	-11.30	-1.3723	-10.52	-1.3320	-10.66	-1.9953	-8.18	1.4780	14.35
(normal)	σ	1.4825	14.98	1.4735	14.64	1.4730	15.34	2.0484	12.25	0.9959	10.38
Depart 10am	μ	0.9778	10.51	0.9734	10.29	0.9944	10.45	1.5333	11.57	0.8365	7.63
(normal)	σ	0.8369	7.69	0.8428	7.73	0.8334	7.43	0.9425	6.48	0.1142	1.15
Depart 2pm	μ	0.1410	1.59	0.1286	1.35	0.1257	1.32	0.4343	2.78	1.0019	9.42
(normal)	σ	0.9761	9.64	0.9768	9.51	0.9893	10.07	1.3703	9.25	0.3462	3.61
	$\gamma_{Ignored}$	-	-	-	-	-	-	-0.8846	12.85[1]	-	-
	$P_{Ignored}$	-		-		-		29.22%		-	
	$P_{Ignored}$ 95% C.I.	-	-	-	-	-	-	14.14%	50.85%	-	-
Virgin Blue	μ	0.1829	3.85	0.1827	3.84	0.1829	3.81	0.1551	3.28	0.5252	5.06
(normal)	σ	0.3458	6.48	0.3392	6.36	0.3460	6.22	0.3443	6.40	-0.8839	-6.30
Air NZ	μ	-0.4524	-8.44	-0.4501	-8.39	-0.4568	-8.46	-0.4643	-8.41	0.5110	6.53
(normal)	σ	0.4372	9.08	0.4329	8.98	0.4360	9.03	0.4590	8.52	-3.2343	2.99
	$\gamma_{Ignored}$	-	-	-	-	-	-	-	-	0.1173	15.39[1]
	$P_{Ignored}$	-		-		-		-		52.93%	
	$P_{Ignored}$ 95% C.I.	-	-	-	-	-	-	-	-	31.49%	73.34%
Model fits											
LL		-2307.35		-2306.78		-2306.40		-2286.10		-2298.53	
Parameters		14		15		14		15		15	
ρ^2		0.4999		0.5001		0.5001		0.5045		0.5019	
Adjusted ρ^2		0.4983		0.4983		0.4985		0.5028		0.5001	
AIC		1.1054		1.1056		1.1050		1.0958		1.1017	

1. t-ratio for difference to zero percent ANA

is present between the Rayleigh distribution and the ANA mass, with the former capturing some of the ANA.

The RPANA1 flight time 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 ($7.09; \chi^2_{1,.05} = 3.84$), but not the Rayleigh ($1.53; \chi^2_{1,.05} = 3.84$). 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 does not have improved model fit over the RPL model.

The fourth model in Table 4 models nonattendance to departure time only. Normally distributed departure time parameters give the best performance for both the RPL and RPANA models. 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 rates recovered by the ANA models exceed the top end of the range, at 53.42 and 52.17 percent.

The final model in Table 4 models nonattendance to airline only. 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 rates under the ANA models of 78.98 and 80.72 percent exceed the upper end of the confidence interval.

From the evidence presented thus far, it appears as if the RPANA model is limited in this study to handling ANA to departure time and airline. Two techniques will be introduced that capture nonattendance to fare and flight time *and* lead to the RPL model being outperformed. The first approach, detailed in the next section, achieves this by introducing stated ANA as covariates in the ANA assignment models. The second approach is to allow ANA to fare and flight time to be correlated, by including both in the same ANA assignment model. A RPCANA2 model was estimated, with ANA modelled for fare and flight time only. On the AIC, this RPCANA2 model (1.1030) outperforms the equivalent RPIANA model (1.1050), the best fitting RPIANA model (lognormal fare and Rayleigh flight time at 1.1049), and the best RPL model tested (uniform fare and normal flight time at 1.1044). Full details of this model are not reported, but fare and flight time are correlated in both of the RPANA4 models presented in Section 4.3.4.

4.3.3 Covariates in ANA

All of the RPANA models estimated thus far have treated the probability of attribute nonattendance as being the same across respondents. However, introducing covariates into the ANA assignment models allows these probabilities to vary across respondents. The use of covariates is motivated by the potential to improve model fit, outperform the RPL model for fare and flight time, and leverage and gain insights into 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.

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 3.17 percent.

The flight time model could be estimated, and strongly outperformed the baseline model without the covariate. Table 5 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 4, 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.

Table 5: RPANA1 models with stated ANA as a covariate for modelled ANA

		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	
	$\gamma_{Ignored}$	-2.9011	8.97	-	-
	$P_{Ignored\|StatedAttended}$	5.21%		-	
	$\theta_{StatedIgnored}$	2.9735	10.44	-	-
	$P_{Ignored\|StatedIgnored}$	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
	$\gamma_{Ignored}$	-	-	-1.3335	19.20
	$P_{Ignored\|StatedAttended}$	-		20.86%	
	$\theta_{StatedIgnored}$	-	-	2.3311	13.40
	$P_{Ignored\|StatedIgnored}$	-		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	
Parameters		15		16	
ρ^2		0.5018		0.5068	
Adjusted ρ^2		0.5000		0.5049	
AIC		1.1019		1.0914	

The implied ANA rates to flight time for both stated attenders and stated ignorers are telling. Only a small proportion of stated attenders, 5.21 percent, ignore flight time, suggesting that stated attendance is fairly accurate. In contrast, only 51.81 percent of stated ignorers 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 the coefficient to zero for these respondents is untenable. However, stated ANA is a source of information that can be used to improve RPANA model fit.

Table 5 also documents a RPANA1 model that leverages stated nonattendance to departure time. The covariate model strongly outperforms the baseline model documented in Table 4 ($20.34; \chi^2_{1,.05} = 3.84$). 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. 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.

4.3.4 Multiple attribute nonattendance

Another possible reason for the lack of improvement in model fit when moving from the RPL to the RPANA1 fare and flight time models 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, with the incidence rates of all *combinations* of fare and flight time nonattendance estimated. Section 4.3.2 concluded by noting that such a model did result in the RPANA model outperforming the best RPL model.

A series of RPANA2 models were first estimated to gain further insight into whether ANA is independent across the attributes. For each pair of attributes, a RPIANA2 and a RPCANA2 model were estimated. Comparisons were performed on model fits and the ANA rates for each attendance pattern across the two attributes. The only decisive evidence for correlation in ANA was found between fare and flight time. Consequently, a model is presented here that models nonattendance to all attributes, with three ANA assignment models, one each for departure time, airline, and the combination of fare and flight time. During estimation, the incidence rate of the combination of ignored fare and attended flight time approached zero ($\gamma_{ac} \to -\infty$), and so this class was dropped from the ANA assignment model. This RPANA4 model is the first presented in Table 6.

This model nests the RPANA1 departure time and airline models from Table 4, as well as the aforementioned RPANA2 model with correlated ANA to fare and flight time. Log likelihood ratio tests reveal that all three previous models are outperformed by the current model. The ANA rate for fare, 4.05 percent, is higher than the rate of 2.12 percent with the RPIANA1 fare model with the same lognormal fare distribution, which hints that a false assumption of independence of ANA might bias the recovered ANA rates. Comparing ANA rates between this model and the three nested models reveals a fair degree of consistency. The notable exception is 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 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 and the model outputs. In this dataset, the series of RPANA2 models

Table 6: RPANA4 and RPCANA4 models

				RPANA4			**RPCANA4**		
				Param.	***t*-ratio**		**Param.**	***t*-ratio**	
Fare		μ		-1.9112	-29.21		-1.8917	-26.58	
(lognormal)		σ		0.7600	9.77		0.7293	9.86	
Flight time		μ		-0.0834	-14.64		-0.0840	-15.05	
(const. $\triangle$)		σ		-	-		-	-	
Depart 6am		μ		-1.9702	-9.73		-1.9548	-7.91	
(normal)		σ		1.9985	12.39		1.9923	12.42	
Depart 10am		μ		1.4931	9.48		1.4997	10.58	
(normal)		σ		0.9475	6.16		0.9466	6.59	
Depart 2pm		μ		0.3643	2.69		0.3439	2.25	
(normal)		σ		1.3563	10.20		1.3587	8.75	
Virgin Blue		μ		0.2730	3.64		0.2957	3.03	
(normal)		σ		0.4666	6.41		0.4699	4.88	
Air NZ		μ		-0.7926	-5.92		-0.8432	-6.28	
(normal)		σ		0.5074	5.59		0.5218	6.19	
Fare	**Flight time**	**Dep. time**	**Airline**	**Param.**	**s.e.**	**Rate**	**Param.**	**s.e.**	**Rate**
Ignore	Ignore	Ignore	Ignore	-	-	0.49%	-3.0596	0.8919	1.60%
Attend	Ignore	Ignore	Ignore	-	-	0.54%	-	-	-
Attend	Attend	Ignore	Ignore	-	-	11.05%	-1.0069	0.4836	12.49%
Ignore	Ignore	Attend	Ignore	-	-	1.30%	-	-	-
Attend	Ignore	Attend	Ignore	-	-	1.45%	-	-	-
Attend	Attend	Attend	Ignore	-	-	29.46%	0.0190	0.4677	34.83%
Ignore	Ignore	Ignore	Attend	-	-	0.62%	-3.2006	0.9016	1.39%
Attend	Ignore	Ignore	Attend	-	-	0.68%	-3.0610	0.9609	1.60%
Attend	Attend	Ignore	Attend	-	-	13.90%	-1.3173	0.6390	9.15%
Ignore	Ignore	Attend	Attend	-	-	1.64%	-2.8500	0.7140	1.98%
Attend	Ignore	Attend	Attend	-	-	1.82%	-2.5082	1.6017	2.78%
Attend	Attend	Attend	Attend	-	-	37.04%	-	-	34.18%
Ignore	Ignore	-	-	-3.1174	0.5041	4.05%	-	-	4.97%
Attend	Ignore	-	-	-3.0116	0.7031	4.50%	-	-	4.38%
Attend	Attend	-	-	-	-	91.45%	-	-	90.64%
Ignore	-	-	-	-	-	4.05%	-	-	4.97%
Attend	-	-	-	-	-	95.95%	-	-	95.03%
-	Ignore	-	-	-	-	8.55%	-	-	9.36%
-	Attend	-	-	-	-	91.45%	-	-	90.64%
-	-	Ignore	-	-0.9805	0.2614	27.28%	-	-	26.24%
-	-	Attend	-	-	-	72.72%	-	-	73.76%
-	-	-	Ignore	-0.2290	0.5705	44.30%	-	-	48.92%
-	-	-	Attend	-	-	55.70%	-	-	51.08%
Model fits									
LL				-2275.35			-2272.11		
Parameters				17			21		
ρ^2				0.5069			0.5076		
Adjusted ρ^2				0.5049			0.5051		
AIC				1.0916			1.0920		

estimated suggest that this is not the case, and that correlation is present only for fare and flight time. 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.

Initially, all 16 ANA combinations were modelled. 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, consistent with the RPANA4 model just presented. The log likelihood became flat, and the standard errors for the ANA assignment parameters associated with these combinations became extremely large. Three more ANA combinations were dropped, because their incidence rate approached zero. The final specification modelled nine combinations of ANA, requiring eight parameters in the ANA assignment model. Table 6 details the model results. While the log likelihood is better than the RPANA4 model 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 RPCANA4 model presented here is inferior on the AIC (1.0920 verses 1.0916).

The ANA rates above the dashed line are the estimated rates for each retained 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. Rates for each ANA combination for the RPANA4 model are also reported, and were arrived at by multiplying the appropriate attendance or nonattendance probability for each attribute (or combination of attributes, for fare and flight time). 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.

In sum, the RPANA4 model is more appealing than the RPCANA4. Model fit is slightly better on the AIC, and the model is more parsimonious. Further, the RPANA4 model allows ANA covariates such as stated ANA to be entered more directly against the attribute itself.

5. Discussion and conclusion

The RPANA model can be evaluated in terms of its econometric performance, and its behavioural appeal. Econometrically, the model performs well. In the empirical context presented in this paper, it outperforms the RPL model, and the two forms of ANA model that exist in the literature. Unlike the ANA model, it can capture preference heterogeneity in addition to ANA. Further, it provides a way to overcome the problem of random parameter distributions having some coefficients with implausible sign, since this may stem from attribute nonattendance (Hensher, 2007), and these coefficients may be approximating ANA. Unlike the method proposed by Hess and Hensher (2010), the RPANA model doesn't require the choice of an arbitrary threshold value, or require sequential estimation. A specific case of the RPANA model was developed by Hess et al. (2012), however the RPANA model developed in this paper is more general, and can range from capturing full correlation in ANA, to as much independence in ANA as the data will support.

The RPANA model has appeal in the context of travel behaviour, as attribute nonattendance is plausible in a range of circumstances. This paper identifies considerable ANA in short haul flights to departure time and airline, moderate ANA to flight time, and low levels of ANA to fare. The extra product features of long haul flights, such as in-flight entertainment and differing seat pitches, are also likely to appeal to only part of the population. It is possible, though, that ANA is not actually capturing true preferences or processing rules. The attribute may have no influence on choice, and in effect not be attended to, because the attribute levels may not be sufficiently differentiated, or may be in an inappropriate range (Hensher et al., 2012a). If these levels and tradeoffs are reflective of choice scenarios that would be encountered in real life, either now or in some plausible future scenario, then ANA that is induced in this way is reasonable, and capturing it may indeed be preferable. For example, some drivers may not attend to a toll, cordon charge, or road user charge for any values that would plausibly be introduced, or variable time-of-day charging may be of insufficient range over time to have salience. However, if the ANA is the consequence of poor experimental design in a stated choice study, then while the RPANA model may do a good job of identifying and isolating ANA, the ANA might have limited behavioural meaning. Care must be taken when generating the experimental design. Careful piloting might help prevent such problems.

This paper identifies a number of important considerations when implementing the RPANA model. The risk of identification problems cannot be ignored. Forcing random parameter distributions associated with linearly coded attributes to be constrained in sign appears to be a necessary but not sufficient condition for identification. Such a constraint was adequate for most parameters and distributions tested. Distributions that span zero can be introduced when multiple parameters are specified for an attribute, however it is necessary to effects code the attribute. Of the distributions that are constrained in sign, the lognormal is the most widespread and probably has the most appeal. However, other distributions should also be tested. Beyond the obvious motivation to find the best distribution that fits the data, differences in the ANA rate recovered as the distribution varied in this paper's empirical application suggest that some degree of confounding between ANA and preference heterogeneity may remain. Which distribution leads to the least confounding is difficult to determine, but model fit is still likely a good guide. The ability to assume independence of ANA in the generalised RPANA model is appealing on the grounds of parsimony, but must be used with caution. If there really is correlation in ANA, then the assumption of independence may inhibit any improvement in model fit, and the ANA rates may be biased. The flexibility of the RPANA model introduced in this paper allows ANA to some attributes to be correlated, whilst remaining independent for other attributes. Finally, introducing covariates into the ANA assignment models potentially offers improvements in model fit and increased insight.

References

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.

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.

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.

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.

Collins, A. T., Rose, J. M., Hess, S., 2012. Interactive stated choice surveys: a study of air travel behaviour. Transportation 39, 55–79.

Gilbride, T. J., Allenby, G. M., Brazell, J. D., 2006. Models for heterogeneous variable selection. Journal of Marketing Research 43 (3), 420–430.

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., 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., 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., 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.

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. 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., Stathopoulos, A., Campbell, D., O'Neill, V., Caussade, S., 2012. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. Transportation.

Hole, A. R., 2011. A discrete choice model with endogenous attribute attendance. Economics Letters 110 (3), 203–205.

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.

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.

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.