

A question of taste: recognising the role of latent preferences and attitudes in analysing food choices

Vikki O’Neill^a Stephane Hess^{*b} Danny Campbell^c

^aMedical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, CB2 0SR,
vikki.oneill@mrc-bsu.cam.ac.uk

^bInstitute for Transport Studies, University of Leeds, Leeds, LS2 9JT, s.hess@its.leeds.ac.uk

^cEconomics Division, Stirling Management School, University of Stirling, Stirling, FK9 4LA,
danny.campbell@stir.ac.uk

Abstract

There has long been substantial interest in understanding consumer food choices, where a key complexity in this context is the potentially large amount of heterogeneity in tastes across individual consumers, as well as the role of underlying attitudes towards food and cooking. The present paper underlines that both tastes and attitudes are unobserved, and makes the case for a latent variable treatment of these components. Using empirical data collected in Northern Ireland as part of a wider study to elicit intra-household trade-offs between home-cooked meal options, we show how these latent sensitivities and attitudes drive both the choice behaviour as well as the answers to supplementary questions. We find significant heterogeneity across respondents in these underlying factors and show how incorporating them in our models leads to important insights into preferences.

Keywords: food preferences; latent variables; stated choice; taste heterogeneity

1 Introduction

There has long been interest in better understanding consumers’ food choices, with a focus on people’s motivations, preferences and habits. Recently, particular emphasis has been put on eating habits within an obesity risk context.

Food choices are complex as well as frequent. In a recent study, [Wansink and Sobal \(2007\)](#) estimated that a person can make over 200 food and beverage related decisions every day. [Asp](#)

*Corresponding author. Tel.: +44 (0)113 34 36611.

(1999) in turn discusses in detail some of the factors which affect consumers when they are deciding what to eat, particularly cultural, psychological and lifestyle factors as well as food trends to name but a few. Work by Lennernäs et al. (1997) has highlighted the role of quality/freshness, price, taste, as well as family preferences and trying to eat healthily, while Drewnowski and Darmon (2005) consider the effects of taste, convenience and economic constraints on food choices. Lennernäs et al. (1997) also found that respondents in different socio-economic categories select different factors as contributing a large portion of influence on their food choices. The extent of heterogeneity in preferences is also highlighted in other work. For example, Logue and Smith (1986) indicate that women have higher preferences for low-calorie foods than men and Rappoport et al. (1993) found that insofar as the *health* value of food was concerned, men had a much simpler cognitive structure than women. Consumer information and market research companies are continually developing classification systems which aim to identify different consumer segments and consequently try to predict consumer behaviour (Asp, 1999). These systems make use of important lifestyle factors to describe how consumers make food decisions. With the exception of examples such as above, most food studies focus on a limited socio-geographic based population (Glanz et al., 1998; Jaeger and Meiselman, 2004; Marshall and Bell, 2004).

A large body of work has looked at respondent reported measures of importance of key attributes. For example, Glanz et al. (1998) examine the self-reported importance of taste, nutrition, cost, convenience, and weight control on personal dietary choices and whether these factors vary across demographic groups, are associated with lifestyle choices related to health, and actually predict eating behaviour. They found that the importance placed on taste, nutrition, cost, convenience, and weight control helped predict types of food consumed. A share of studies which have investigated adult preferences for a variety of foods have involved the respondent rating individual food items on either a nine, five or four point scale, wherein the studies reported the mean rating for each food item (see, for example Bell and Marshall, 2003, Drewnowski and Hann, 1999, Jaeger and Meiselman, 2004 and Rappoport et al., 1993).

Whilst simple rating methods can provide rich information about specific food preferences, they do not examine food preference *patterns* which would help elicit more general food preferences. For example, a person's preference for one type of food could be a predictive indicator of that person's preference for another type of food (Logue and Smith, 1986). Across a number of fields, mathemat-

57 ical structures belonging to the family of random utility models have established themselves as the
58 preferred method for the study of choice behaviour at the disaggregate level (Train, 2009). These
59 models quantify the relative importance of the different attributes describing each alternative and
60 are used across fields as diverse as transport, marketing and health economics. This study adds to
61 a growing literature that has used these models to examine food choices and preferences for food
62 attributes (see, for example Campbell and Doherty, 2013, Carlsson et al., 2007, Hu et al., 2004,
63 Jaeger and Rose, 2008, Jaeger et al., 2008, Lusk and Briggeman, 2009, Ortega et al., 2011 and
64 Rigby et al., 2009). More specifically, this paper contributes to the literature where these models
65 have been used to investigate the link between food choice, diet and health (e.g., Balcombe et al.,
66 2010; Gracia et al., 2009; Mueller Loose et al., 2013).

67 The present paper illustrates how advanced choice models can be used to obtain a better
68 understanding of consumer food choices. In particular, we recognise, in line with previous work,
69 that there exist significant differences in preferences across individual consumers. We hypothesise
70 that while some of these differences can be linked to socio-demographic characteristics, others
71 cannot. The standard modelling approach for such “unexplained” differences would be a model
72 allowing for random taste heterogeneity. Any information about sensitivities¹ and differences in
73 sensitivities would be inferred solely on the basis of the choices made by respondents. We use a
74 more refined approach that allows us to make use of the supplementary information provided by
75 respondents in ranking questions and attitudinal questions within a hybrid choice model making
76 use of latent variables (e.g., Ben-Akiva et al., 2002a,b; Bolduc et al., 2005). This gives us a better
77 understanding of what drives food choices, and the differences in these drivers across the population.

78 The remainder of this paper is organised as follows. Section 2 presents an overview of the
79 empirical data and methods used in this study. This is followed in Section 3 by a discussion of the
80 results for both the base models and the latent variable models. Finally, a concluding discussion is
81 presented in Section 4.

¹ We have chosen to use the term ‘sensitivities’ here, as we felt it more appropriate in this specific context, as the more commonly used term ‘preferences’ can be seen to relate to alternatives, not just attributes.

82 2 Material and methods

83 2.1 Survey work

84 Data were collected as part of a wider study to elicit intra-household trade-offs between home-
85 cooked meal options. The respondents used for the survey formed a random sample of Northern
86 Ireland households, and face-to-face interviews were used for preference elicitation.

87 Table 1 shows the socio-demographic characteristics of the respondents. Just over a third of the
88 respondents were aged between 35 and 50, with the rest split evenly above and below these ages.
89 The average income per week was £211, with 48% of the respondents in full-time employment.
90 10% had at least a degree level education.

91 2.1.1 Stated choice component

92 In the stated choice component of the survey, respondents were presented with the choice between
93 three different meal options representing a typical evening meal that they would share with their
94 partner at home. After a qualitative stage, including consultation with experts and assisted in-
95 terviews with respondents, we conducted a pilot study. Following this, we were able to select the
96 following attributes to describe the meal options: calories, cooking time, food type and cost. Taste
97 was not included as a direct variable in the choice tasks as it would be subject to *interpretation* by
98 the respondent. Instead, “food type” was used as a proxy for taste. Three levels were used for each
99 attribute, where the specific combinations presented in a given choice scenario were obtained from
100 a D-efficient experimental design with Bayesian priors (Bliemer and Rose, 2010; Rose and Bliemer,
101 2009), produced using NGene (ChoiceMetrics, 2012). A D-efficient design was chosen so as to min-
102 imise the asymptotic variance covariance matrix. The final design contained 24 rows which were
103 divided into 3 blocks of 8 choices, where each respondent was asked to complete 8 choice tasks. To
104 ensure that any heterogeneity retrieved in both the parameter estimates as well as the variances of
105 the error terms is not simply an artefact of the design of choice set scenarios (Arentze et al., 2003),
106 we used orthogonal blocking, and randomly assigned people to blocks.

107 Table 2 shows the three levels used for the different attributes, where “Cost” represented the
108 total cost for all of the ingredients needed to produce a typical evening meal, which would feed
109 both the respondent and his or her partner. To allow respondents to better relate to the attribute

Tab. 1: Socio-demographic characteristics

Age		Female		Male		Total			
18-24		32	11%	27	9%	59	10%		
25-34		71	24%	66	23%	137	23%		
35-50		100	34%	100	34%	200	34%		
51-59		35	12%	40	14%	75	13%		
60-64		22	8%	20	7%	42	7%		
65-75		32	11%	35	12%	67	11%		
75+		0	0%	4	1%	4	1%		
Income									
Per week		Per Year		Female		Male		Total	
Less than £150	Less than £7,800	142	49%	91	31%	233	40%		
£150 - £299	£7,800 - £15,599	98	34%	121	41%	219	38%		
£300 - £449	£15,600 - £23,399	41	14%	59	20%	100	17%		
£450 - £599	£23,400 - £31,199	8	3%	15	5%	23	4%		
£600+	£31,200+	3	1%	6	2%	9	2%		
Employment									
		Female		Male		Total			
In full-time employment		109	37%	174	60%	283	48%		
In part-time employment		68	23%	18	6%	86	15%		
Self-employed		7	2%	11	4%	18	3%		
Unemployed		36	12%	30	10%	66	11%		
Retired		48	16%	50	17%	98	17%		
Student/Otherwise not working		24	8%	9	3%	33	6%		
Education									
		Female		Male		Total			
No qualifications		52	18%	46	16%	98	17%		
CSE/GCSE/O Levels		148	51%	141	48%	289	49%		
A Level/Baccalaureate		46	16%	36	12%	82	14%		
Vocational Qualification		18	6%	38	13%	56	10%		
Degree		25	9%	25	9%	50	9%		
Postgraduate Degree		3	1%	6	2%	9	2%		
Total		292	100%	292	100%	584	100%		

110 levels for calories, cooking time and food type, they were provided with illustrative reference cards
111 that showed what type of meal could be expected for given attribute combinations. We chose cost
112 levels of £5, £10 and £15 pounds after conducting a pilot study; the large cost differences were
113 found to be needed as respondents were reacting very strongly to the different levels of the other
114 attributes, causing the cost attribute to become insignificant when smaller price differences were
115 used.

116 In each choice task, respondents were asked to choose their most preferred option for a typical
117 evening meal that they would share together with their partner at home, and which would be

Tab. 2: Attribute levels

Attribute	Levels
Calories (<i>per portion</i>)	Less than 400 calories Between 400 and 600 calories Over 600 calories
Cooking Time	Less than 30 minutes Between 31 and 60 minutes Over 60 minutes
Food Type (<i>proxy for taste</i>)	Asian Italian Local
Cost	£5 £10 £15

118 cooked at home. An example choice scenario is shown in Figure 1. We decided against explicitly
 119 including a “no choice” option, but if a respondent could not decide, then this was recorded as
 120 a “Don’t know” by the interviewer². For the present study, we made use of responses from 584
 121 individuals, giving 4,672 observations in total.

122 2.1.2 Supplementary questions

123 In addition to completing the choice tasks, respondents were also asked to state their most preferred
 124 and least preferred level of each of the three non-cost attributes. A summary of the information
 125 obtained in this manner is shown in Figure 2, where the first two columns in each subfigure show
 126 the responses to the questions eliciting the respondent’s *most* preferred options, for females and
 127 males respectively, and the last two columns in each subfigure show the responses to the questions
 128 eliciting the respondent’s *least* preferred options, for females and males respectively.

129 The results from this exercise are in line with expectations and the prior literature. We can see
 130 that for calories, 49% of the interviewed women prefer the medium calories range, with a total of
 131 80% preferring fewer than 600 calories in their meal. Whilst this preference pattern is also shown
 132 by male respondents, the level of uncertainty (“Don’t know”) is increased, especially for the least

² We acknowledge this potential limitation within the data (Olsen and Swait, 1997), but this approach was taken as the sample size was quite small and we did not want to reduce the data further by encouraging “Don’t know” responses. However, although respondents were not told upfront that they could state “Don’t know”, if they did so, it was recorded. Further, if the respondent stated “Don’t know” at any point in the questionnaire and it was recorded down then they would know that it was safe to say “Don’t know”, meaning that only the first instance of “Don’t know” could be subject to any bias.

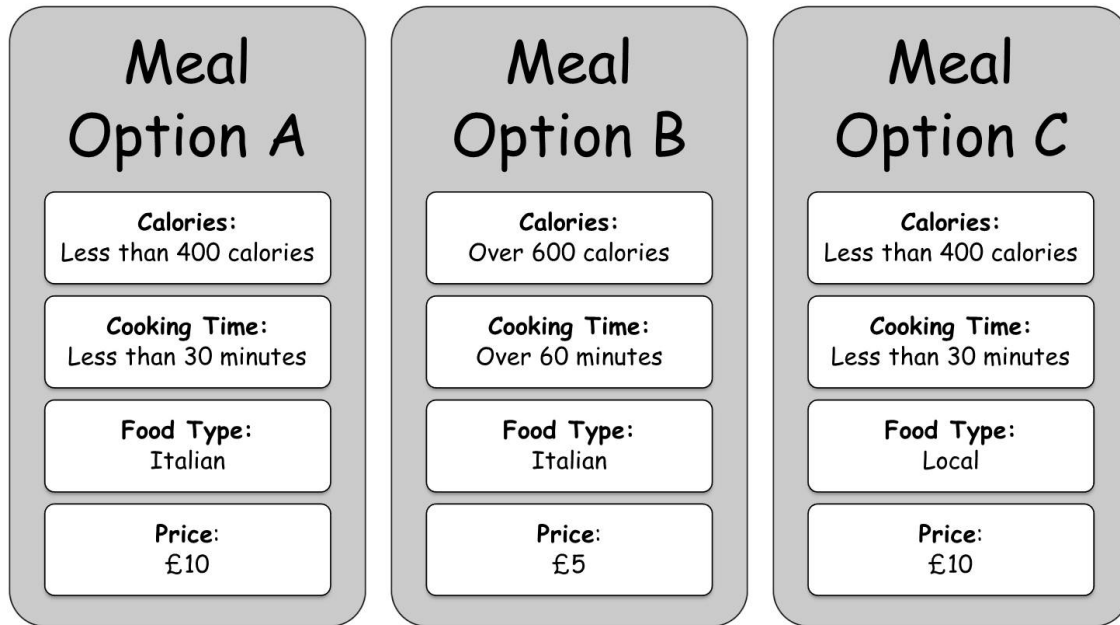


Fig. 1: Example choice task

133 preferred calorie level. With regards to cooking time, medium cooking time is again the most
 134 preferred, while high cooking time is generally the least preferred. Overall, the question which
 135 encountered the fewest “Don’t know” responses was that which asked respondents for their most
 136 preferred food types. Local food was the most popular choice; this is in line with findings by
 137 [McIlveen and Chestnutt \(1999\)](#), where they conclude that greater product awareness needs to be
 138 instigated by retailers in Northern Ireland in order to inform consumers of the larger range of food
 139 products available to them and consequently encourage greater uptake. [McIlveen and Chestnutt](#)
 140 [\(1999\)](#) found that the Italian food sector represented a growth area, whereas Indian and other newly
 141 developing food sectors were not yet evident in Northern Ireland. Note that this relates to cooking
 142 meals at home rather than eating out, where there is an abundance of international restaurants
 143 available.

144 As a final component, respondents were also presented with three questions relating to attitudes
 145 towards cooking. In particular, respondents were asked to indicate their level of agreement (on a
 146 five-point Likert scale) with three statements, namely:

- 147 • “Cooking is not much fun”;
- 148 • “Compared with other daily decisions, my food choices are not very important”; and

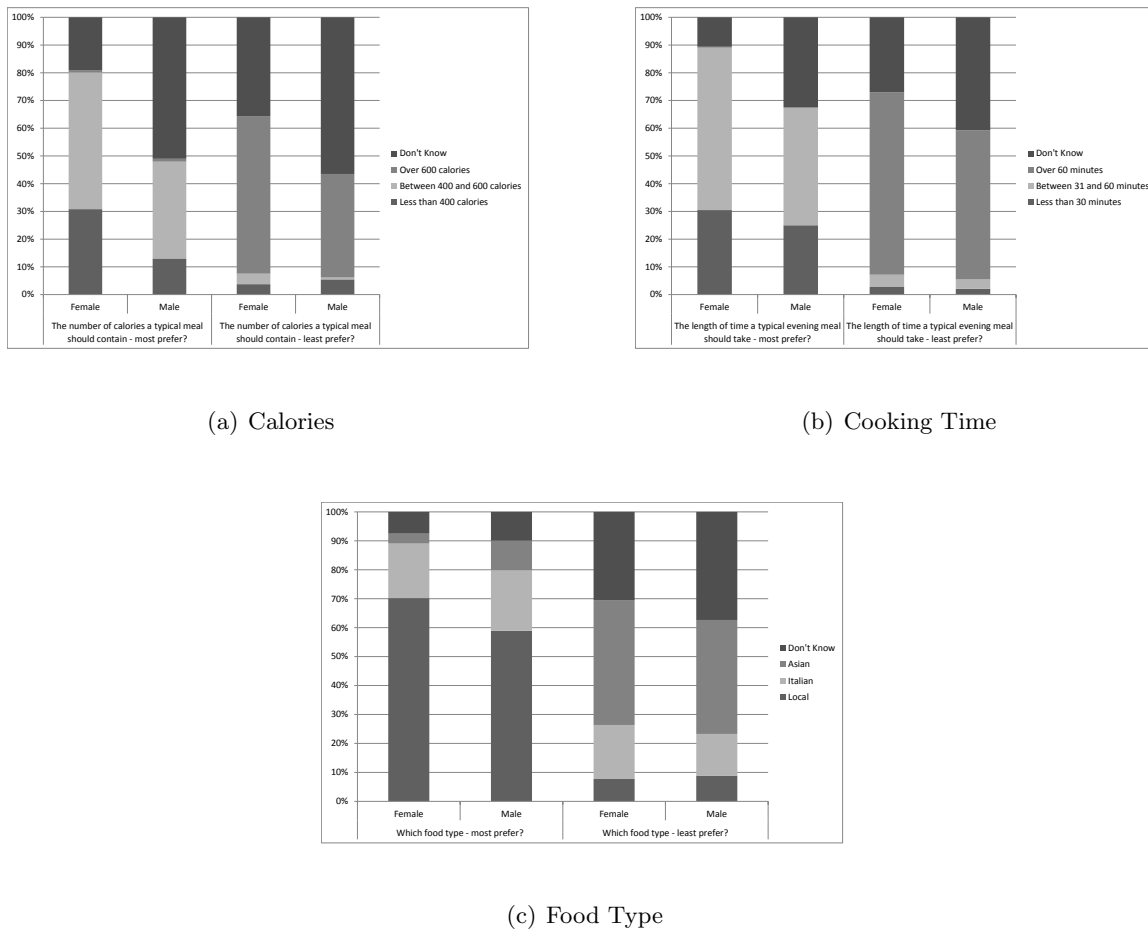


Fig. 2: Attribute importance rankings

149 • “*I enjoy cooking for others and myself*”.

150 Figure 3 shows a summary of the responses to the three attitudinal questions, highlighting a more
 151 positive attitude towards cooking for female respondents, along with a higher prevalence of “Don’t
 152 know” responses for male respondents.

153 The inclusion of these statements was driven in part by the success achieved in [Bell and Marshall](#)
 154 (2003) and [Marshall and Bell](#) (2004) at being able to classify differences in food choices and food
 155 choice patterns by using a measure of food involvement, namely the “Food Involvement Scale”
 156 (FIS). [Bell and Marshall](#) (2003) define food involvement as ‘the level of importance of food in a
 157 person’s life’. They also assume that as a result of this, the level of food involvement will vary
 158 across individuals. [Bell and Marshall](#) (2003) and [Laaksonen](#) (1994, pg. 8-9) suggest that food
 159 involvement is a mediating variable, acting between stimulus objects and response, depending on

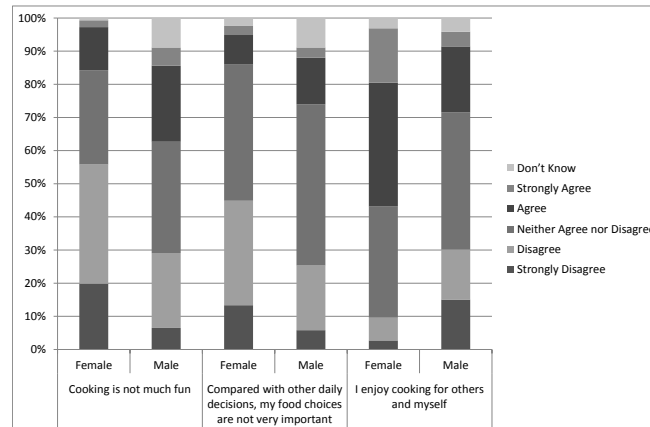


Fig. 3: Answers to attitudinal questions relating to cooking

160 both the characteristics of the stimulus object and those of the consumer.

161 2.2 Base model specification

162 As a first step, we estimate simple Multinomial Logit (MNL) models on our data, where we use
 163 the panel specification of the sandwich estimator to recognise the repeated choice nature of the
 164 data in the computation of standard errors (cf. [Daly and Hess, 2011](#)). All models reported in this
 165 paper were coded in Ox 6.2 ([Doornik, 2007](#)). For the MNL model, we used maximum likelihood
 166 estimation, while maximum simulated likelihood estimation was used for the hybrid models, with
 167 simultaneous estimation of all model components.

168 Two different specifications are used. In the first model, the deterministic component of utility³
 169 for respondent n and alternative i in choice task t (out of 8) is written as:

³ In the MNL specification, the random component of the utility function follows a type I extreme value distribution.

$$\begin{aligned}
V_{int} = & \beta_{\text{LowCal}} \text{LowCal}_{int} + \beta_{\text{HighCal}} \text{HighCal}_{int} + \\
& \beta_{\text{LowTime}} \text{LowTime}_{int} + \beta_{\text{HighTime}} \text{HighTime}_{int} + \\
& \beta_{\text{Asian}} \text{Asian}_{int} + \beta_{\text{Italian}} \text{Italian}_{int} + \\
& \beta_{\text{Cost}} \text{Cost}_{int} \qquad \qquad \qquad \forall 1 \leq i \leq 3 \qquad (1)
\end{aligned}$$

$$V_{4nt} = \delta_{\text{DK}} \text{DK}_{4nt}, \qquad (2)$$

170 where, as an example, LowCal_{int} is set to 1 if alternative i has the low calories level (and is set to
171 0 if alternative i has a calories level other than low), and where β_{LowCal} is the associated marginal
172 utility coefficient, which is to be estimated. Equation 1 shows the utility individual n will receive
173 if they select any of the first three alternatives, whereas Equation 2 shows the utility individual n
174 will receive through the selection of the “Don’t know” option (displayed as alternative 4, in this
175 case)⁴. Other than cost, the attributes were entered as dummy variables in order to allow us to
176 capture any non-linear preference structure for these attributes, where the middle level was used
177 as the base (i.e. sensitivity fixed to zero).

178 The specification thus far has assumed that the sensitivities to the different attribute levels (i.e.
179 the preferences) are constant across individuals in our sample. To address this shortcoming, we
180 make use of a revised specification that allows for differences in sensitivities for the three non-cost
181 attributes by age group as well as by gender. For each level (other than middle), we thus estimate
182 a base coefficient, along with offsets for male respondents, respondents under the age of 35 and
183 respondents over the age of 50, using the middle age group as the base. This specification is shown
184 in Equation 3, where, for example, $\Delta_{\text{Italian};\text{Male}}$ shows the shift in the utility for Italian food for a
185 male respondent aged 35-49 years relative to a female respondent aged 35-49 years.

$$\begin{aligned}
V_{int} = & \beta_{\text{LowCal};\text{Base}} \text{LowCal}_{int} + \Delta_{\text{LowCal};\text{Male}} \text{LowCal}_{int} \\
& + \Delta_{\text{LowCal};\text{Under 35}} \text{LowCal}_{int} + \Delta_{\text{LowCal};\text{Over 50}} \text{LowCal}_{int}
\end{aligned}$$

⁴ We previously tested for left-to-right bias by estimating alternative specific constants for $i - 1$ of the hypothetical choices and found none, so we decided to use an alternative specific constant for the “Don’t know” choices.

$$\begin{aligned}
& +\beta_{\text{HighCal};\text{Base}}\text{HighCal}_{int} + \Delta_{\text{HighCal};\text{Male}}\text{HighCal}_{int} \\
& +\Delta_{\text{HighCal};\text{Under 35}}\text{HighCal}_{int} + \Delta_{\text{HighCal};\text{Over 50}}\text{HighCal}_{int} \\
& +\beta_{\text{LowTime};\text{Base}}\text{LowTime}_{int} + \Delta_{\text{LowTime};\text{Male}}\text{LowTime}_{int} \\
& +\Delta_{\text{LowTime};\text{Under 35}}\text{LowTime}_{int} + \Delta_{\text{LowTime};\text{Over 50}}\text{LowTime}_{int} \\
& +\beta_{\text{HighTime};\text{Base}}\text{HighTime}_{int} + \Delta_{\text{HighTime};\text{Male}}\text{HighTime}_{int} \\
& +\Delta_{\text{HighTime};\text{Under 35}}\text{HighTime}_{int} + \Delta_{\text{HighTime};\text{Over 50}}\text{HighTime}_{int} \\
& +\beta_{\text{Asian};\text{Base}}\text{Asian}_{int} + \Delta_{\text{Asian};\text{Male}}\text{Asian}_{int} \\
& +\Delta_{\text{Asian};\text{Under 35}}\text{Asian}_{int} + \Delta_{\text{Asian};\text{Over 50}}\text{Asian}_{int} \\
& +\beta_{\text{Italian};\text{Base}}\text{Italian}_{int} + \Delta_{\text{Italian};\text{Male}}\text{Italian}_{int} \\
& +\Delta_{\text{Italian};\text{Under 35}}\text{Italian}_{int} + \Delta_{\text{Italian};\text{Over 50}}\text{Italian}_{int} \\
& +\beta_{\text{Cost}}\text{Cost}_{int} \qquad \qquad \qquad \forall 1 \leq i \leq 3 \qquad (3)
\end{aligned}$$

2.3 Integrated Choice and Latent Variable (ICLV) model specification

The base model with deterministic heterogeneity allows for variations in sensitivities as a function of age and gender. However, it is easily conceivable that additional differences exist which cannot entirely be linked to socio-demographic characteristics. Rather than relying on a simple random coefficients specification, we propose to make use of the additional information collected from respondents in terms of attribute rankings as well as attitudinal questions. Specifically, we hypothesise that these additional data can serve as proxies for the underlying differences in sensitivities. However, it is important to recognise that answers to attribute ranking questions and attitudinal questions do not provide us with a direct error-free measure of the actual underlying sensitivities. Indeed, they are merely a function of these sensitivities. Similarly, these data points are likely to be correlated with other unobserved effects, and their incorporation as explanatory variables in our choice models would thus put us at risk of endogeneity bias.

To allow us to use the additional data while not exposing ourselves to the risk of measurement error and endogeneity bias, we make use of a hybrid model specification in which the answers to ranking questions and attitudinal questions are treated as dependent rather than explanatory variables. A number of latent variables are then used to create a link between a given respondent's

202 choices and his/her answers to these additional questions. Within such an Integrated Choice and
 203 Latent Variable (ICLV) model, the responses to the subjective questions are modelled jointly with
 204 the actual choice processes, all the while maintaining the assumption that *both* processes are at
 205 least in part influenced by the latent attitudes. This approach integrates choice models with
 206 latent variable models resulting in an improvement in the understanding of preferences and allow
 207 us to make use of additional data sources. The theoretical developments of such hybrid choice
 208 models centre on the work of Ben-Akiva et al. (2002a,b) and Bolduc et al. (2005), with numerous
 209 applications, for example Abou-Zeid et al. (2010), Alvarez-Daziano and Bolduc (2009), Daly et al.
 210 (2012a), Fosgerau and Bjørner (2006), Hess and Beharry-Borg (2012), Johansson et al. (2006) and
 211 Yáñez et al. (2010).

212 Our work makes use of seven latent variables:

- 213 • two latent variables linked to the underlying sensitivities to the low and high levels for calories,
 214 α_{LowCal} and α_{HighCal} ;
- 215 • two latent variables linked to the underlying sensitivities to the low and high levels for cooking
 216 time, α_{LowTime} and α_{HighTime} ;
- 217 • two latent variables linked to the underlying sensitivities to Italian and Asian food, α_{Italian}
 218 and α_{Asian} ; and
- 219 • one latent variable linked to general attitudes towards food, hereafter known as the ‘*cooking*’
 220 attitude, α_{Cooking} .

221 We use a linear in attributes specification for the deterministic part, and write:

$$\alpha_{k,n} = \gamma_{\alpha_k} z_n + \eta_{k,n},$$

$$k = \text{LowCal, HighCal, LowTime, HighTime, Italian, Asian, Cooking} \quad (4)$$

222 where $\gamma_{\alpha_k} z_n$ represents the deterministic part of $\alpha_{k,n}$, with, z_n being a vector of socio-demographic
 223 variables, γ_{α_k} being a vector of estimated parameters and $\eta_{k,n}$ being a random disturbance, which
 224 follows a standard Normal distribution across respondents.

225 Hereafter, α_n represents the vector of latent attitudes for respondent n . These latent variables
 226 are now used as explanatory variables in the utility function, which is rewritten as:

$$V_{int} = f(\beta, x_{int}, \delta, \alpha_n, \tau) \quad (5)$$

227 where τ is a vector of parameters that explain the impact of the vector of latent variables α_n on
 228 the utility of alternative i , possibly in interaction with the attributes x_{int} and the parameters β .

229 At the same time, we use the latent variables to explain the responses to the ranking questions
 230 and the attitudinal questions. In particular, the first two latent variables, α_{LowCal} and α_{HighCal} , are
 231 used to explain the ranking of the three different calorie levels, the following two latent variables,
 232 α_{LowTime} and α_{HighTime} , are used for the ranking of the three different time levels, and the fifth and
 233 sixth latent variables, α_{Italian} and α_{Asian} , are used to explain the ranking of the three different food
 234 types. Finally, the seventh latent variable, α_{Cooking} , is used to explain the answers to the three
 235 attitudinal questions about cooking.

236 For each of the three non-cost attributes, respondents were asked to state their most preferred
 237 and least preferred level (i.e. *best* and *worst* level respectively). We represent the underlying
 238 sensitivities to the different levels in a utility framework, where, for the example of the calories
 239 attribute, we have that:

- 240 • the utility for *low* calories is given by the latent variable for the underlying sensitivity to low
 241 calories, i.e. α_{LowCal} , plus a parameter $\mu_{R,\text{LowCal}}$; where $\mu_{R,\text{LowCal}}$ captures the mean ranking
 242 in the sample;
- 243 • the utility for *high* calories is given by the latent variable for the underlying sensitivity to
 244 high calories, i.e. α_{HighCal} , plus a parameter $\mu_{R,\text{HighCal}}$; where $\mu_{R,\text{HighCal}}$ captures the mean
 245 ranking in the sample; and
- 246 • the utility for *medium* calories is set to zero.

247 For the response to the *worst* attribute level, the sign of the utilities was reversed⁵. Respondents

⁵ Clearly, the actual latent variable used in the two specifications needs to be the same here, so the only assumption relates to using the same μ_R terms in the best and worst (with sign change) specifications. We found no significant asymmetry in these terms, hence our decision. The same does not apply for the “Don’t know” term where separate constants were used.

248 were also allowed to opt out of each ranking question, by giving a “Don’t know” response to either
 249 their best or worst preferred level. The utilities for such responses are given by constants, where
 250 separate constants are used for the best and worst rankings, given the differential rates of “Don’t
 251 know”.

252 The actual probabilities for the observed responses to the best and worst ranking questions are
 253 now given by:

$$P_{\text{cal-best},n} = \frac{\mathbb{I}_{\text{LC},n}^B e^{\mu_{R,\text{LowCal}} + \alpha_{\text{LowCal},n}} + \mathbb{I}_{\text{MC},n}^B + \mathbb{I}_{\text{HC},n}^B e^{\mu_{R,\text{HighCal}} + \alpha_{\text{HighCal},n}} + \mathbb{I}_{\text{DK BC},n}^B e^{\delta_{R,\text{DK BestCal}}}}{e^{\mu_{R,\text{LowCal}} + \alpha_{\text{LowCal},n}} + 1 + e^{\mu_{R,\text{HighCal}} + \alpha_{\text{HighCal},n}} + e^{\delta_{R,\text{DK BestCal}}}} \quad (6)$$

$$P_{\text{cal-worst},n} = \frac{\mathbb{I}_{\text{LC},n}^W e^{-\mu_{R,\text{LowCal}} - \alpha_{\text{LowCal},n}} + \mathbb{I}_{\text{MC},n}^W + \mathbb{I}_{\text{HC},n}^W e^{-\mu_{R,\text{HighCal}} - \alpha_{\text{HighCal},n}} + \mathbb{I}_{\text{DK WC},n}^W e^{\delta_{R,\text{DK WorstCal}}}}{e^{-\mu_{R,\text{LowCal}} - \alpha_{\text{LowCal},n}} + 1 + e^{-\mu_{R,\text{HighCal}} - \alpha_{\text{HighCal},n}} + e^{\delta_{R,\text{DK WorstCal}}}} \quad (7)$$

254 where:

- 255 • $\mathbb{I}_{\text{LC},n}^B$ is an indicator variable, equal to 1 if respondent n choose ‘Low’ as his/her most preferred
 256 calorie level and 0 otherwise;
- 257 • $\mathbb{I}_{\text{MC},n}^B$ is an indicator variable, equal to 1 if respondent n choose ‘Medium’ as his/her most
 258 preferred calorie level and 0 otherwise;
- 259 • $\mathbb{I}_{\text{HC},n}^B$ is an indicator variable, equal to 1 if respondent n choose ‘High’ as his/her most preferred
 260 calorie level and 0 otherwise; and
- 261 • $\mathbb{I}_{\text{DK BC},n}^B$ is an indicator variable, equal to 1 if respondent n did not know his/her most
 262 preferred calorie level and 0 otherwise.

263 Equivalently \mathbb{I}^W is an indicator variable for the least favourite rankings. The parameters $\delta_{R,\text{DK BestCal}}$
 264 and $\delta_{R,\text{DK WorstCal}}$ give the utility for the “Don’t know” choices.

265 A corresponding specification was used for the ranking questions for time and food type. From
 266 this, we then obtain:

$$L(R_n | \alpha_{*,n}) = P_{\text{cal-best},n} P_{\text{cal-worst},n} P_{\text{time-best},n} P_{\text{time-worst},n} P_{\text{type-best},n} P_{\text{type-worst},n}; \quad (8)$$

267 which gives the probability of observing the specific responses given by respondent n to the ranking
 268 questions as a product of logit probabilities which is conditional on the first six latent variables,
 269 where $\alpha_{*,n} = \langle \alpha_{LowCal,n}, \alpha_{HighCal,n}, \alpha_{LowTime,n}, \alpha_{HighTime,n}, \alpha_{Italian,n}, \alpha_{Asian,n} \rangle$.

270 The specification used for the cooking indicators is somewhat different. In line with [Daly et al.](#)
 271 (2012a), we treat the responses to these three attitudinal questions using an ordered logit model
 272 specification (see also [Bierlaire, 2008](#)). The probability of observing a given value s for the k^{th}
 273 indicator (with $k = 1, 2, 3$) for respondent n , with $s = 1, \dots, 5$, where $s = 1$ indicates a strong
 274 agreement with the statement and $s = 5$ indicates a strong disagreement, is now given by:

$$P(I_{k,n} | \alpha_{Cooking,n}) = \frac{e^{\psi_{k,s} - \zeta_{I_k} \alpha_{Cooking,n}}}{1 + e^{\psi_{k,s} - \zeta_{I_k} \alpha_{Cooking,n}}} - \frac{e^{\psi_{k,s-1} - \zeta_{I_k} \alpha_{Cooking,n}}}{1 + e^{\psi_{k,s-1} - \zeta_{I_k} \alpha_{Cooking,n}}} \quad (9)$$

275 where the estimated effect of the latent variable $\alpha_{Cooking,n}$ on this indicator is given by ζ_{I_k} , and the
 276 probability of the actual observed response is then given by:

$$L(I_{k,n} | \alpha_{Cooking,n}) = \sum_{s=1}^S \mathbb{I}_s^{k,n} \left[\frac{e^{\psi_{k,s} - \zeta_{I_k} \alpha_{Cooking,n}}}{1 + e^{\psi_{k,s} - \zeta_{I_k} \alpha_{Cooking,n}}} - \frac{e^{\psi_{k,s-1} - \zeta_{I_k} \alpha_{Cooking,n}}}{1 + e^{\psi_{k,s-1} - \zeta_{I_k} \alpha_{Cooking,n}}} \right] \quad (10)$$

277 where $\mathbb{I}_1^{k,n} = 1$ if respondent n gives level 1 as the answer to the k^{th} attitudinal question, and zero
 278 otherwise. For normalisation, we set $\psi_{k,0} = -\infty$ and $\psi_{k,5} = +\infty$ and estimate the four intermediate
 279 thresholds, where $\psi_{k,s} \geq \psi_{k,s-1}$. Finally, we set $L(I_n | \alpha_{Cooking,n}) = \prod_{k=1}^3 L(I_{k,n} | \alpha_{Cooking,n})$.

280 Our joint model now has three components in the likelihood function; a choice model, a mea-
 281 surement model for the ranking questions, and a measurement model for the three attitudinal
 282 questions. These are driven by structural equations for utilities and latent variables, respectively.
 283 The likelihood for the observed sequence of choices for respondent n is given by $L(y_n | \beta, \delta, \tau, \alpha_n)$,
 284 which is a product of logit probabilities, and a function of the parameters of the base choice model
 285 (grouped together into β), the τ parameters and the vector of seven latent variables α . The likeli-
 286 hood for the measurement model for the ranking question is given by $L(R_n | \mu_R, \delta, \alpha_{*,n})$ which is
 287 a function of the first six latent variables as well as a set of constants and the mean ranking pa-
 288 rameters. Finally, the likelihood for the measurement model for the attitudinal questions is given

289 by $L(I_n | \zeta_I, \psi, \alpha_{\text{Cooking},n})$, which is a function of the ζ terms, the threshold parameters ψ , and the
 290 seventh latent variable.

291 In combination, the log-likelihood function is thus given by:

$$LL(\beta, \gamma, \tau, \zeta_I, \psi, \mu_R, \delta) = \sum_{n=1}^N \ln \int_{\eta} L(y_n | \cdot) L(I_n | \cdot) L(R_n | \cdot) g(\eta) d\eta \quad (11)$$

292 Equation 11 is dependent on the latent variables, which is shown by the integration over η , the
 293 random component of α , and the fact that the log-likelihood is a function of γ , which drives the
 294 deterministic part of α . Hence, in addition to the parameters estimated for the standard model,
 295 the estimation of this model entails the estimation of the vector of τ terms, the parameters of
 296 the various measurement equations, and the socio-demographic interaction terms γ . As previously
 297 mentioned, maximum simulated likelihood estimation was used for this model in the absence of a
 298 closed form solution for the log-likelihood function in Equation 11.

299 The entire structure of the model is represented graphically in Figure 4. At the top of the graph,
 300 we have the indicators, I_k ; “Calorie Ranking”, “Time Ranking”, “Food Type Ranking” and “Cook-
 301 ing Attitudes” (for which we have three indicator functions). These indicators are explained using
 302 the seven latent variables, which in turn are a function of socio-demographic variables (in addition
 303 to having a random component). The latent variables are then at the same time interacted with
 304 the coefficients of the choice model (β), which are possibly also interacted with socio-demographic
 305 indicators, and which, in interaction with the attribute levels, explain the choices observed in the
 306 data.

307 Before proceeding with the discussion of results, it should of course be acknowledged that the
 308 use of ICLV leads to increased estimation cost and the need for datasets to contain additional
 309 indicators, but this is commonly the case. Additionally, there is the added demand for the analyst
 310 to specify structural equations for the latent variables and to make decisions relating to functional
 311 form, including for the measurement model. However, when done in a competent manner, the
 312 advantages can be very substantial, where, as explained previously, as key advantage of ICLV over
 313 more standard models (e.g. mixed logit and latent class) is its ability to use additional data to
 314 explain the heterogeneity across decision makers, and to provide further insights.

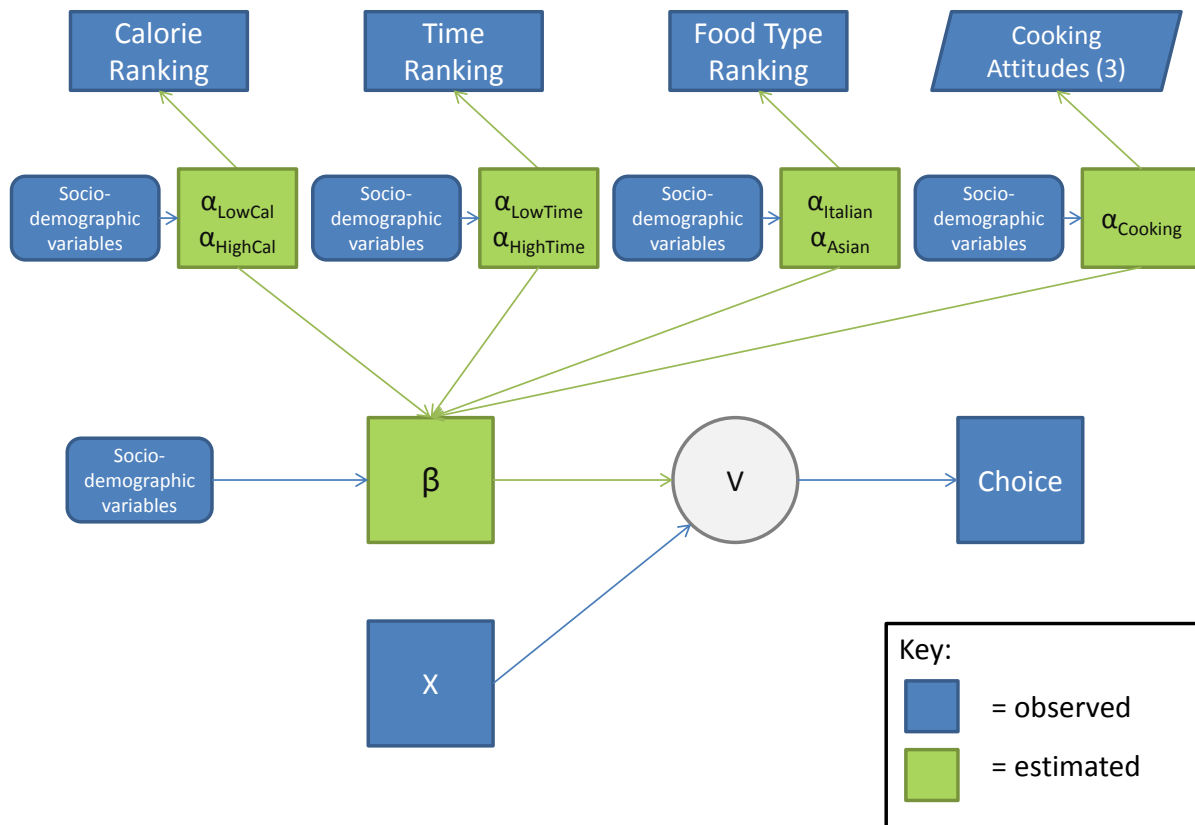


Fig. 4: ICLV model outline

315 3 Results

316 3.1 Base model results

317 The results for the two base models are summarised in Table 3. Looking first at the model without
 318 socio-demographic interactions, we can see that the coefficients for low calories (β_{LowCal}) is positive
 319 and significant while the coefficient for high time ($\beta_{HighTime}$) is negative and significant. This
 320 indicates that low levels of calories are preferred to medium levels of calories, while medium time
 321 is preferred to high time. The signs for the coefficients for high calories ($\beta_{HighCal}$) and low time
 322 ($\beta_{LowTime}$) are not in line with this, but the coefficients are not statistically significant, making
 323 the sign irrelevant and showing that there is no difference from the sensitivity for the medium
 324 level in these cases; at the aggregate level, the respondents are not distinguishing between high
 325 calories and the base level medium calories, or between low time and the base level of medium

Tab. 3: Base MNL model and MNL with age and gender effects

	Base MNL		MNL with age and gender	
	est.	rob. <i>t</i> -rat.	est.	rob. <i>t</i> -rat.
$\beta_{\text{LowCal;Base}}$	0.2468	4.74	0.5050	4.97
$\Delta_{\text{LowCal;Male}}$	-	-	-0.1970	-2.00
$\Delta_{\text{LowCal;Under 35}}$	-	-	-0.3231	-2.66
$\Delta_{\text{LowCal;Over 50}}$	-	-	-0.1652	-1.36
$\beta_{\text{HighCal;Base}}$	0.0341	0.69	0.0341	0.35
$\Delta_{\text{HighCal;Male}}$	-	-	0.0310	0.33
$\Delta_{\text{HighCal;Under 35}}$	-	-	0.1261	1.08
$\Delta_{\text{HighCal;Over 50}}$	-	-	-0.1826	-1.56
$\beta_{\text{LowTime;Base}}$	-0.0142	-0.34	0.1048	1.22
$\Delta_{\text{LowTime;Male}}$	-	-	-0.0061	-0.07
$\Delta_{\text{LowTime;Under 35}}$	-	-	-0.1402	-1.28
$\Delta_{\text{LowTime;Over 50}}$	-	-	-0.2086	-2.00
$\beta_{\text{HighTime;Base}}$	-0.2197	-6.52	-0.1220	-1.57
$\Delta_{\text{HighTime;Male}}$	-	-	-0.0319	-0.45
$\Delta_{\text{HighTime;Under 35}}$	-	-	-0.2219	-2.42
$\Delta_{\text{HighTime;Over 50}}$	-	-	-0.0182	-0.21
$\beta_{\text{Italian;Base}}$	-0.0599	-1.20	0.1852	2.00
$\Delta_{\text{Italian;Male}}$	-	-	-0.0357	-0.37
$\Delta_{\text{Italian;Under 35}}$	-	-	-0.2900	-2.57
$\Delta_{\text{Italian;Over 50}}$	-	-	-0.4213	-3.34
$\beta_{\text{Asian;Base}}$	-0.3275	-6.65	-0.0888	-0.95
$\Delta_{\text{Asian;Male}}$	-	-	0.0247	0.26
$\Delta_{\text{Asian;Under 35}}$	-	-	-0.5272	-4.62
$\Delta_{\text{Asian;Over 50}}$	-	-	-0.2605	-2.12
β_{Cost}	-0.0493	-7.92	-0.0504	-8.07
δ_{DK}	-3.8274	-20.87	-3.8540	-20.97
\mathcal{LL}	-5,192.85		-5,141.8	

time. We can also see that, as expected, the coefficients for Italian (β_{Italian}) and Asian (β_{Asian}) food are negative, meaning that respondents prefer the base of Local food to these alternatives, albeit that the difference with Italian food is not statistically significant. The cost coefficient (β_{Cost}) has the expected negative estimate, while the strong negative estimate for the constant for the “Don’t know” alternative (δ_{DK}) reflects the low rate of respondents indicating indecision between alternatives.

Turning to the model incorporating socio-demographic interactions, using a likelihood ratio test, we obtain an improvement in log-likelihood by 51.85 units over the base model at the cost of 18 additional parameters - this is highly significant giving a likelihood-ratio test value of 103.7

335 compared to a χ^2_{18} critical value of 34.81 at the 99% level. While we note a significant negative
 336 shift in preferences towards low calories for males, we do not find significant differences between
 337 males and females for any of the other attributes, a finding which is contrary to much of the food
 338 preference literature. On the other hand, we observe a number of significant age interactions.
 339 Notably, we observe a lower preference for low calorie levels for respondents under the age of 35,
 340 along with reduced preferences (or increased dislike) of high time as well as Italian and Asian food.
 341 For respondents over 50 years of age, we note a significant negative shift in preferences for low time,
 342 as well as once again Italian and Asian food.

343 3.2 Integrated Choice and Latent Variable (ICLV) model results

344 The specification for our latent variable model made use of the base specification from the MNL
 345 model without socio-demographic interactions, given that these are now dealt with in the latent
 346 variable specification.

347 In the choice model, the first six latent variables were interacted with the associated parameter,
 348 e.g. the latent variable for low calories was interacted with the β parameter for low calories. The
 349 latent variable for general cooking attitude was interacted with all non-cost coefficients in the choice
 350 model, with the exception of high time where no meaningful effect was retrieved. With this in mind,
 351 we have that the utilities for the first three alternatives are now given as:

$$\begin{aligned}
 V_{int} = & \beta_{\text{LowCal}} \text{LowCal}_{int} + \tau_{\alpha_{\text{LowCal}}, \beta_{\text{LowCal}}} \alpha_{\text{LowCal}, n} + \tau_{\alpha_{\text{Cooking}}, \beta_{\text{LowCal}}} \alpha_{\text{Cooking}, n} \\
 & + \beta_{\text{HighCal}} \text{HighCal}_{int} + \tau_{\alpha_{\text{HighCal}}, \beta_{\text{HighCal}}} \alpha_{\text{HighCal}, n} + \tau_{\alpha_{\text{Cooking}}, \beta_{\text{HighCal}}} \alpha_{\text{Cooking}, n} \\
 & + \beta_{\text{LowTime}} \text{LowTime}_{int} + \tau_{\alpha_{\text{LowTime}}, \beta_{\text{LowTime}}} \alpha_{\text{LowTime}, n} + \tau_{\alpha_{\text{Cooking}}, \beta_{\text{LowTime}}} \alpha_{\text{Cooking}, n} \\
 & + \beta_{\text{HighTime}} \text{HighTime}_{int} + \tau_{\alpha_{\text{HighTime}}, \beta_{\text{HighTime}}} \alpha_{\text{HighTime}, n} \\
 & + \beta_{\text{Italian}} \text{Italian}_{int} + \tau_{\alpha_{\text{Italian}}, \beta_{\text{Italian}}} \alpha_{\text{Italian}, n} + \tau_{\alpha_{\text{Cooking}}, \beta_{\text{Italian}}} \alpha_{\text{Cooking}, n} \\
 & + \beta_{\text{Asian}} \text{Asian}_{int} + \tau_{\alpha_{\text{Asian}}, \beta_{\text{Asian}}} \alpha_{\text{Asian}, n} + \tau_{\alpha_{\text{Cooking}}, \beta_{\text{Asian}}} \alpha_{\text{Cooking}, n} \\
 & + \beta_{\text{Cost}} \text{Cost}_{int}
 \end{aligned} \tag{12}$$

352 while the utility for alternative 4 remains the same as in the MNL models.

Tab. 4: Estimation results for choice model component

	est.	rob. <i>t</i> -rat.
β_{LowCal}	0.4103	4.57
β_{HighCal}	-0.2388	-2.79
β_{LowTime}	0.0258	0.42
β_{HighTime}	-0.2444	-6.38
β_{Italian}	0.0444	0.55
β_{Asian}	-0.3197	-3.19
β_{Cost}	-0.0532	-7.55
δ_{DK}	-3.9231	-20.61
$\tau_{\alpha_{\text{LowCal}}, \beta_{\text{LowCal}}}$	0.6740	7.50
$\tau_{\alpha_{\text{HighCal}}, \beta_{\text{HighCal}}}$	0.3783	2.78
$\tau_{\alpha_{\text{LowTime}}, \beta_{\text{LowTime}}}$	0.6065	7.78
$\tau_{\alpha_{\text{HighTime}}, \beta_{\text{HighTime}}}$	0.0303	0.75
$\tau_{\alpha_{\text{Italian}}, \beta_{\text{Italian}}}$	0.3187	5.53
$\tau_{\alpha_{\text{Asian}}, \beta_{\text{Asian}}}$	0.6476	6.80
$\tau_{\alpha_{\text{Cooking}}, \beta_{\text{LowCal}}}$	-0.2089	-3.04
$\tau_{\alpha_{\text{Cooking}}, \beta_{\text{HighCal}}}$	0.0779	1.21
$\tau_{\alpha_{\text{Cooking}}, \beta_{\text{LowTime}}}$	-0.0519	-1.17
$\tau_{\alpha_{\text{Cooking}}, \beta_{\text{Italian}}}$	-0.0707	-1.21
$\tau_{\alpha_{\text{Cooking}}, \beta_{\text{Asian}}}$	-0.0080	-0.12
<i>Choice component</i> \mathcal{LL}	-5,044.01	
<i>Overall</i> \mathcal{LL}	-10,666.60	

353 The specification of the measurement equations is as discussed in Section 2.3. The means
354 of the latent variables were set to zero, and an extensive amount of testing was conducted to
355 establish significant socio-demographic interactions, focussing on age and gender, where only the
356 most significant interactions were retained, as discussed later in this section.

357 The estimation results for the choice model component, as outlined in Equation 12 above, are
358 shown in Table 4. The overall fit for the hybrid model, also shown in Table 4, cannot be directly
359 compared to that for the MNL model as it jointly models the choices and responses to attitudinal
360 and ranking questions (c.f. Equation 11). However, it is possible to factor out the component of
361 the log-likelihood relating to the choice model, conditional on the other components. This gives
362 us a log-likelihood of $-5,044.01$, which shows that the model offers a better statistical fit for the
363 choice data compared to the two base models, but no formal statistical tests are conducted, given
364 the conditioning on other model components. Extensive discussions on this issue are given in Vij
365 and Walker (forthcoming).

Tab. 5: Estimation results for structural equation model for latent attitudes

Latent variable	Estimated parameter	est.	rob. t -rat.
α_{LowCal}	$\gamma_{\text{LowCal}_{<35}}$	-0.2594	-1.95
α_{HighCal}	$\gamma_{\text{HighCal}_{\text{Male}}}$	0.5171	2.08
	$\gamma_{\text{HighCal}_{<35}}$	0.5011	3.03
α_{LowTime}	$\gamma_{\text{LowTime}_{50+}}$	-0.2595	-1.85
α_{HighTime}	$\gamma_{\text{HighTime}_{\text{Male}}}$	0.5171	2.56
	$\gamma_{\text{Italian}_{\text{Male}}}$	0.3186	1.76
α_{Italian}	$\gamma_{\text{Italian}_{<35}}$	-0.5442	-2.54
	$\gamma_{\text{Italian}_{50+}}$	-0.9269	-4.24
	$\gamma_{\text{Asian}_{\text{Male}}}$	0.2087	1.39
α_{Asian}	$\gamma_{\text{Asian}_{<35}}$	-0.5072	-2.99
	$\gamma_{\text{Asian}_{50+}}$	-0.3310	-1.86
	$\gamma_{\text{Cooking}_{\text{Male}}}$	0.6713	5.98
α_{Cooking}	$\gamma_{\text{Cooking}_{<35}}$	0.5018	3.67
	$\gamma_{\text{Cooking}_{50+}}$	0.2534	1.80

366 We first observe that β_{HighCal} has changed in sign and has also become significant compared
367 with the base model. This is in line with the preferences found above in Figure 2. Two additional
368 parameters, namely β_{LowTime} and β_{Italian} , also undergo sign changes, but the coefficients remain
369 insignificant. For the first six latent variable effects, we can see that, in line with expectations, a
370 higher value for the underlying attribute sensitivity leads to a more positive parameter in the choice
371 model, albeit that this is not statistically significant for high time. For the final latent variable, i.e.
372 the general cooking attitude, only one effect is significant, indicating that a higher value for the
373 latent attitude equates to a less positive value for the associated low calorie coefficient. As we will
374 see later, this latent variable in fact equates to an *anti-cooking* attitude, meaning that respondents
375 who have a more positive attitude towards cooking also prefer cooking lower calorie meals.

376 As a next step, we look at the structural equations for the seven latent variables, as outlined
377 above in Equation 4, with estimates summarised in Table 5. These results show that male respon-
378 dents have a more positive value for the latent variables for high calories, high time and Italian and
379 Asian food types. The result for high time may seem counter-intuitive, but a possible explanation
380 could be that whilst they would prefer to have meals that take longer to cook, they do not neces-
381 sarily want to be responsible for creating the meal. We also see that male respondents have a more
382 positive value for the general latent cooking attitude, where it is important to remember that this
383 is in fact an *anti-cooking* attitude, which explains the sign. The same applies for the low and high

Tab. 6: Estimation results for measurement models for rankings of attributes; Calories, Cooking Time and Food Type

	est.	rob. <i>t</i> -rat.
<i>Calories: α_{LowCal} and α_{HighCal}</i>		
$\mu_{R,\text{LowCal}}$	-0.7629	-5.54
$\mu_{R,\text{HighCal}}$	-4.0481	-15.30
$\delta_{R,\text{DK Most Cal}}$	-0.1595	-1.65
$\delta_{R,\text{DK Least Cal}}$	3.5868	17.00
<i>Cooking Time: α_{LowTime} and α_{HighTime}</i>		
$\mu_{R,\text{LowTime}}$	-0.5965	-4.73
$\mu_{R,\text{HighTime}}$	-4.2649	-16.80
$\delta_{R,\text{DK Most Time}}$	-0.7959	-7.30
$\delta_{R,\text{DK Least Time}}$	3.3050	14.61
<i>Food Type: α_{Italian} and α_{Asian}</i>		
$\mu_{R,\text{Italian}}$	-0.9207	-4.91
$\mu_{R,\text{Asian}}$	-2.1267	-10.59
$\delta_{R,\text{DK Most Type}}$	-1.9328	-12.79
$\delta_{R,\text{DK Least Type}}$	2.0953	13.74

384 age groups. In addition, being under the age of 35 has a negative effect on the latent variable for
 385 low calories, as well as for Italian and Asian food types, but a positive affect on the latent variable
 386 for high calories. Lastly, respondents aged over 50 have a less positive value for the latent variable
 387 for low time, as well as non-local food.

388 As discussed in Section 2.3, the measurement component explains the observed attribute rank-
 389 ings (c.f. Equations 6 and 7) in addition to the answers for the cooking attitudinal questions (c.f.
 390 Equation 9). The results for the measurement model for attribute rankings are summarised in Ta-
 391 ble 6, whereas the results for the three attitudinal questions are shown in Table 7. We will discuss
 392 each of these in turn below.

393 Concerning Table 6, the negative signs for the six mean ranking parameters are a reflection of
 394 the fact that, across attributes, the middle level tended to be ranked highest by respondents. The
 395 signs for the “Don’t know” constants reflect the low rates for choosing “Don’t know” in response
 396 to the *best* level question, and the high rate for choosing it in response to the *worst* level question.
 397 This is an indication that respondents find it harder to evaluate their least preferred option and as
 398 a result, are more inclined to state “Don’t know”.

399 We finally turn to the results for the measurement model for the three attitudinal questions,
 400 which are shown in Table 7. We can see that the thresholds are all increasing in magnitude, as

Tab. 7: Estimation results for measurement model for latent attitude to Cooking, α_{Cooking}

	est.	rob. <i>t</i> -rat.
<i>Cooking is not much fun</i>		
$\zeta_{\text{Cooking 1}}$	3.1146	7.13
Threshold 1: $\psi_{1,1}$	-2.2387	-4.84
Threshold 2: $\psi_{1,2}$	1.3287	2.88
Threshold 3: $\psi_{1,3}$	4.7295	7.00
Threshold 4: $\psi_{1,4}$	8.3355	8.82
<i>Compared with other daily decisions, my food choices are not very important</i>		
$\zeta_{\text{Cooking 2}}$	1.6174	8.51
Threshold 1: $\psi_{2,1}$	-2.1674	-8.41
Threshold 2: $\psi_{2,2}$	0.2199	0.88
Threshold 3: $\psi_{2,3}$	3.4837	9.70
Threshold 4: $\psi_{2,4}$	5.6278	12.32
<i>I enjoy cooking for others and myself</i>		
$\zeta_{\text{Cooking 3}}$	-2.8201	-8.87
Threshold 1: $\psi_{3,1}$	-6.2423	-9.38
Threshold 2: $\psi_{3,2}$	-4.6090	-8.10
Threshold 3: $\psi_{3,3}$	-0.8788	-2.21
Threshold 4: $\psi_{3,4}$	2.6166	5.76

401 is required by the model. Additionally, we see positive estimates for the effect in the first two
402 equations, and a negative effect in the third model. This means that a more positive value for
403 the seventh latent variable leads to stronger agreement with the statements that “*Cooking is not*
404 *much fun*” and “*Compared with other daily decisions, my food choices are not very important*”,
405 but increased disagreement with the statement that “*I enjoy cooking for others and myself*”. This
406 is in line with an interpretation of this latent variable as an *anti-cooking* attitude, which explains
407 the role of this latent variable in the choice model as well as the signs of the socio-demographic
408 interactions in its structural equation.

409 3.3 WTP / Marginal Rates of Substitution

410 As a final step, we turn our attention to implied willingness to pay (WTP) patterns and other
411 marginal rates of substitution.

412 We first look at the WTP patterns from our base MNL model without socio-demographic
413 interactions, shown in Table 8(a). The context of the survey was a study of home-cooked meal
414 options, namely respondents’ preferences for a typical evening meal that they would share with

Tab. 8: Willingness to pay (WTP) measures

(a) Base MNL model:	
	WTP
LowCal	5.00
HighCal	0.69
LowTime	-0.29
HighTime	-4.45
Italian	-1.21
Asian	-6.64

(b) MNL with age and gender effects:										
	Percentiles									
	5	10	25	50	75	90	95	Mean	SD	
LowCal	-0.30	-0.30	2.83	3.61	6.74	10.01	10.01	4.85	3.25	
HighCal	-2.94	-2.94	-2.33	1.29	3.18	3.79	3.79	0.66	2.50	
LowTime	-2.18	-2.18	-2.06	-0.70	1.96	2.08	2.08	-0.25	1.73	
HighTime	-7.45	-7.45	-6.82	-3.42	-2.78	-2.42	-2.42	-4.33	2.01	
Italian	-5.39	-5.39	-4.68	-2.08	2.97	3.67	3.67	-1.30	3.52	
Asian	-12.22	-12.22	-11.73	-6.44	-1.76	-1.27	-1.27	-6.69	4.32	

(c) ICLV Model:										
	Percentiles									
	5	10	25	50	75	90	95	Mean	SD	
LowCal	-17.96	-13.05	-4.77	4.30	13.47	21.64	26.53	4.31	13.52	
HighCal	-13.48	-10.67	-5.91	-0.61	4.70	9.48	12.33	-0.60	7.84	
LowTime	-20.03	-15.82	-8.82	-1.03	6.75	13.73	17.90	-1.04	11.53	
HighTime	-5.41	-5.20	-4.84	-4.45	-4.05	-3.69	-3.48	-4.45	0.59	
Italian	-12.71	-10.34	-6.36	-1.90	2.59	6.67	9.05	-1.87	6.62	
Asian	-28.77	-24.23	-16.64	-8.20	0.24	7.84	12.41	-8.20	12.51	

415 their partner at home. Consequently, the *cost* element of this represented the total cost for all of
416 the ingredients needed to produce this evening meal which would feed them both. We can thus
417 interpret the willingness to pay (WTP) measures as the extra cost that the respondent would be
418 willing to pay for the evening meal to be shifted away from the middle (base) level (or have to
419 obtain in price reductions to accept such a change). In these results, negative WTP measures
420 reflect the fact that some attribute levels are undesirable when compared to the middle level. For
421 the base model, we note a positive WTP for moving from middle calorie to low calorie meals, while
422 cost reductions are required at the aggregate level to accept a move to high time or Asian food.
423 The remaining WTP measures relate to parameters that were not statistically significant.

424 Table 8(b) and Table 8(c) show the corresponding results for the MNL model with gender and
425 age interactions as well as for the ICLV model. In both cases, we now have variation across respon-
426 dents, where the variation in the MNL model is purely deterministic, as a result of incorporating
427 socio-demographics in the model, while the variation in the ICLV models is driven by both the
428 socio-demographic and random components in the structural equations for the latent variables. In
429 both models, we summarise the heterogeneity by presenting the values for a number of points on
430 the sample level distribution, in the form of percentiles. While the signs and size of the mean
431 WTP measures remain in line with the simple MNL results, most WTP measures now show tails
432 of opposite signs - for example, in Table 8(b) we see that the proportion of people who would have
433 a negative WTP for moving from middle calorie to low calorie meals contains between 10-25% of
434 the sample. This reflects the high degree of heterogeneity in the data, where, for the ICLV model,
435 it is also important to acknowledge the potential impact of the Normal distribution on results. We
436 see that the tails from the distributions in the ICLV model are very long and suggest some very
437 high WTP measures for a small share of respondents. It is important to recognise that the Normal
438 distribution is unbounded and this clearly plays a role in these tails. Of further key importance is
439 the strong retrieved impact that the latent attitudes have on sensitivities, with several of the esti-
440 mated τ parameters exceeding the associated coefficient in absolute value, leading to the resulting
441 high level of heterogeneity. It is worth mentioning in this context that we found no evidence of
442 fully lexicographic behaviour in the data.

443 For other marginal rates of substitution, we focus on a shift from medium calories to low
444 calories, and in particular respondents' willingness to accept a move to high time (from medium

Tab. 9: Marginal rates of substitution (MRS)

(a) Base MNL model:		MRS	
Move to Low Cal and accept High Time		1.12	
Move to Low Cal and accept Asian		0.75	

(b) MNL with age and gender effects:		MRS: Percentiles							Mean	SD
		5	10	25	50	75	90	95		
Move to Low Cal and accept High Time		-0.04	-0.04	0.53	0.83	2.42	4.14	4.14	1.65	1.40
Move to Low Cal and accept Asian		-0.03	-0.03	0.30	0.44	4.80	5.68	5.68	2.07	2.32

(c) ICLV Model:		MRS: Percentiles						
		5	10	25	50	75	90	95
Move to Low Cal and accept High Time		-4.12	-2.97	-1.08	0.97	3.05	4.94	6.09
Move to Low Cal and accept Asian		-5.55	-2.64	-0.73	0.16	1.10	3.03	6.04

time) or Asian food (from local food) in return for such a change. For the simple MNL model, Table 9(a) shows that the desire to shift to low calories is stronger than the desire to avoid a shift from medium time to high time, but is not as strong as the desire to avoid a shift from local food to Asian food. For the model with socio-demographic interactions (cf. Table 9(b)), we see strong heterogeneity, where sign changes are a result of some segments disliking low calories or having a positive preference for High Time or Asian food. While the mean is greater than 1 for both marginal rates of substitution, the medians are both lower than 1. This implies that while some respondents have a very strong preference for a move to low calories, the relative preference for avoiding a move to high time or Asian food is stronger for over fifty percent of respondents. This is also reflected in the results for the ICLV model (cf. Table 9(c)), where the use of the Normal distribution implies that means and standard deviations for the marginal rates of substitution cannot be calculated (c.f. Daly et al., 2012b). The use of the Normal distribution is in this case an inherent component of the ICLV structure. Nevertheless, while moments cannot be calculated, we can of course still report medians and other percentiles, as we do.

4 Discussion

In this paper, we have highlighted the potential benefit of using advanced choice models for studying consumers' food choices. In particular, we have considered the impact that attitudes and underlying preferences can have on the decision making process through the use of a latent variable approach. We started with a simple MNL model which revealed that most of the estimates were in line with expectation, and those that were not were found not to be significant. We also estimated a MNL model with variation in sensitivities by age and gender, producing interesting findings, not least in part due to the significant preference differences found between the age groups used.

As a next step, we illustrated how further differences can be accommodated in a latent variable based hybrid model structure which allows us to make use of additional subjective data on attribute rankings and attitudinal questions. Crucially, this model allows us to use such data without risk of measurement error or endogeneity bias. We formulated a model with seven latent variables and showed how this model provides us with important further insights into behaviour. The latent variables are used to explain both differences in sensitivities in the choice model as well as the

473 responses to attribute ranking questions and attitudinal questions. In this context, a number of
474 interesting socio-demographic interactions were also retrieved.

475 Some potential limitations in this study must be acknowledged. Firstly, our dataset may have
476 been subject to some endogeneity issues between cost and quality, that has been previously found
477 in other food studies (Richards and Padilla, 2009)⁶. In addition, at an earlier stage of this work,
478 feedback from our survey interviewers indicated that people were associating low cooking time with
479 low quality food, whereas people were associating a lengthy cooking time with high inconvenience,
480 which may help to explain the counter-intuitive finding of the preferred cooking time being between
481 31 and 60 minutes. Further, a recent paper by Grisolia et al. (2012) mentions an important
482 element in general food choices; the issue of experienced utility vs. expected utility. This could
483 also be an important confounder in our survey, where the types of foods that the respondents had
484 bought and cooked at home previously could have had a bearing on their current food preferences.
485 Finally, the use of the MNL model without socio-demographic variables inside the ICLV model is a
486 simplification. We took this decision primarily with a view to avoiding using the same limited set
487 of socio-demographic variables in two components of the model (utility specification and structural
488 equations for the latent variable) where we were concerned with confounding.

489 The ICLV model has the key advantage of being a very flexible model, allowing the use of a
490 wide set of different indicators. Future work could make use of other factors such as those related
491 to health risk aversion and weight control problems, which unfortunately were not included in the
492 present survey⁷. We believe that there is wide scope for ICLV applications in a food choices context.
493 Indeed, it is well known that preferences vary extensively across consumers and it is conceivable that
494 a large extent of such heterogeneity relates to underlying convictions, preferences and attitudes.
495 Examples for future areas of application include a focus on topics such as health and diet, ethical
496 food sources, organic food, as well as locally sourced food. A further key advantage of the model is in
497 forecasting. Indeed, once the latent variables have been *calibrated* with the help of the measurement
498 model, this component of the model becomes redundant in forecasting, meaning that indicators are
499 no longer needed, and only choices are predicted. With a sufficiently detailed specification for the
500 structural equations, this would also allow forecasting under hypothetical changes to the make-up

⁶ We thank an anonymous referee for conveying this to us.

⁷ We are grateful to an anonymous referee for having pointed out these and many other things to us.

501 of the population of consumers, for example in relation to age and income.

502 **Acknowledgements**

503 The authors would like to thank Andrew Daly and Amanda Stathopoulos for their suggestions.
504 The authors would also like to acknowledge the input of Hannah McClure in early stages of this
505 research. An earlier version of this paper was presented at the second International Choice Mod-
506 elling Conference (ICMC) which was held in Leeds, July 2011 and the authors are grateful for
507 comments received there which provided insightful suggestions for revisions. We also gratefully
508 acknowledge the financial support for the data collection from the UKCRC Centre of Excellence
509 for Public Health (NI).

510 **References**

- 511 Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C., Hess, S., 2010. Attitudes and value of
512 time heterogeneity. In: Van de Voorde, E., Vanelander, T. (Eds.), *Applied transport economics*
513 - A management and policy perspective. De Boeck Publishing, pp. 523–545.
- 514 Alvarez-Daziano, R., Bolduc, D., 2009. Canadian consumers' perceptual and attitudinal responses
515 toward green automobile technologies: an application of hybrid choice models, paper presented
516 at the 2009 EAERE-FEEM-VIU European Summer School in Resources and Environmental
517 Economics: Economics, Transport and Environment, Venice International University, Italy.
- 518 Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses:
519 effects of task complexity, presentation format and literacy. *Transportation Research Part E:*
520 *Logistics and Transportation Review*, 39 (3), 229–244.
- 521 Asp, E. H., 1999. Factors affecting food decisions made by individual consumers. *Food Policy*
522 24 (2/3), 287–294.
- 523 Balcombe, K., Fraser, I., Di Falco, S., 2010. Traffic lights and food choice: a choice experiment
524 examining the relationship between nutritional food labels and price. *Food Policy* 35 (3), 211–220.
- 525 Bell, R., Marshall, D. W., 2003. The construct of food involvement in behavioral research: scale
526 development and validation. *Appetite* 40 (3), 235–244.

- 527 Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-
528 Supan, A., Brownstone, D., Bunch, D. S., Daly, A., de Palma, A., Gopinath, D., Karlstrom,
529 A., Munizaga, M. A., 2002a. Hybrid choice models: progress and challenges. *Marketing Letters*
530 13 (3), 163–175.
- 531 Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., Polydoropoulou, A.,
532 2002b. Integration of choice and latent variable models. In: Mahmassani, H. (Ed.), *In perpetual*
533 *motion: travel behavior research opportunities and challenges*. Pergamon, pp. 431–470.
- 534 Bierlaire, M. 2008. Estimation of discrete choice models with BIOGEME 1.8, biogeme.epfl.ch
- 535 Bliemer, M. C., Rose, J. M., 2010. Construction of experimental designs for mixed logit models
536 allowing for correlation across choice observations. *Transportation Research Part B: Methodolog-*
537 *ical*, 44 (6), 720–734.
- 538 Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid choice models with logit kernel:
539 applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), *Integrated land-use*
540 *and transportation models: behavioural Foundations*. Elsevier, pp. 275–302.
- 541 Campbell, D., Doherty, E., 2013. Combining discrete and continuous mixing distributions to identify
542 niche markets for food. *European Review of Agricultural Economics* 40 (2), 287–312.
- 543 Carlsson, F., Frykblom, P., Lagerkvist, C. J., 2007. Consumer benefits of labels and bans on GM
544 foods: choice experiments with Swedish consumers. *American Journal of Agricultural Economics*
545 89 (1), 152–161.
- 546 ChoiceMetrics, 2012. Ngene 1.1.1 User manual and reference guide, choice-metrics.com
- 547 Daly, A. J., Hess, S., 2011. Simple methods for panel data analysis. Paper presented at the 90th
548 Annual Meeting of the Transportation Research Board, Washington, D.C.
- 549 Daly, A., Hess, S., Patrui, B., Potoglou, D., Rohr, C., 2012a. Using ordered attitudinal indicators
550 in a latent variable choice model: a study of the impact of security on rail travel behaviour.
551 *Transportation* 39 (2), 267–297.

- 552 Daly, A., Hess, S., Train, K., 2012b. Assuring finite moments for willingness to pay estimates from
553 random coefficients models. *Transportation* 39 (1), 19–31.
- 554 Doornik, J. A., 2007. *Object-oriented matrix programming using Ox*, 3rd ed. London: Timberlake
555 Consultants Press and Oxford, doornik.com
- 556 Drewnowski, A., Darmon, N., 2005. Food choices and diet costs: an economic analysis. *The Journal*
557 *of Nutrition* 135 (4), 900–904.
- 558 Drewnowski, A., Hann, C., 1999. Food preferences and reported frequencies of food consumption
559 as predictors of current diet in young women. *The American Journal of Clinical Nutrition* 70 (1),
560 28–36.
- 561 Fosgerau, M., Bjørner, T. B., 2006. Joint models for noise annoyance and willingness to pay for
562 road noise reduction. *Transportation Research Part B: Methodological* 40 (2), 164–178.
- 563 Glanz, K., Basil, M., Maibach, E., Goldberg, J., Snyder, D., 1998. Why americans eat what
564 they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food
565 consumption. *Journal of the American Dietetic Association* 98 (10), 1118–1126.
- 566 Gracia, A., Loureiro, M. L., Nayga, R. M., Jr., 2009. Consumers' valuation of nutritional informa-
567 tion: A choice experiment study. *Food Quality and Preference* 20 (7), 463–471.
- 568 Grisolia, J. M., López, F., Ortúzar, J. D. D., 2012. Sea urchin: From plague to market opportunity.
569 *Food Quality and Preference* 25 (1), 46–56.
- 570 Hensher, D. A., Rose, J. M., 2012. The influence of alternative acceptability, attribute thresh-
571 olds and choice response certainty on automobile purchase preferences. *Journal of Transport*
572 *Economics and Policy*, 46 (3), 451–468.
- 573 Hess, S., Beharry-Borg, N., 2012. Accounting for latent attitudes in willingness to pay studies: the
574 case of coastal water quality improvements in Tobago. *Environmental and Resource Economics*
575 52 (1), 109–131.
- 576 Hu, W., Hünne Meyer, A., Veeman, M., Adamowicz, W., Srivastava, L., 2004. Trading off health,

- 577 environmental and genetic modification attributes in food. *European Review of Agricultural*
578 *Economics* 31 (3) 389–408.
- 579 Jaeger, S. R., Meiselman, H. L., 2004. Perceptions of meal convenience: the case of at-home evening
580 meals. *Appetite* 42 (3), 317–325.
- 581 Jaeger, S. R., Rose, J. M., 2008. Stated choice experimentation, contextual influences and food
582 choice: a case study. *Food Quality and Preference* 19 (6), 539–564.
- 583 Jaeger, S. R., Jørgensen, A. S., Aaslyng, M. D., Bredie, W. L. P., 2008. Best-worst scaling: an
584 introduction and initial comparison with monadic rating for preference elicitation with food
585 products. *Food Quality and Preference* 19 (6), 579–588.
- 586 Johansson, M. V., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on
587 mode choice. *Transportation Research Part A: Policy and Practice* 40 (6), 507–525.
- 588 Laaksonen, P., 1994. *Consumer involvement: concepts and research*. Routledge.
- 589 Lennernäs, M., Fjellström, C., Becker, W., Giachetti, I., Schmitt, A., de Winter, A. R., Kearney, M.,
590 1997. Influences on food choice perceived to be important by nationally-representative samples
591 of adults in the European Union. *European Journal of Clinical Nutrition* 51 (suppl 2), S8–S15.
- 592 Logue, A. W., Smith, M. E., 1986. Predictors of food preferences in adult humans. *Appetite* 7 (2),
593 109–125.
- 594 Lusk, J., Briggeman, B., 2009. Food values. *American Journal of Agricultural Economics* 91 (1)
595 184–196.
- 596 Marshall, D., Bell, R., 2004. Relating the food involvement scale to demographic variables, food
597 choice and other constructs. *Food Quality and Preference* 15 (7/8), 871–879.
- 598 McIlveen, H., Chestnutt, S., 1999. The Northern Ireland retailing environment and its effect on
599 ethnic food consumption. *Nutrition & Food Science* 99 (5), 237–243.
- 600 Mueller Loose, S., Peschel, A., Grebitus, C., 2013. Quantifying effects of convenience and product
601 packaging on consumer preferences and market share of seafood products: The case of oysters.
602 *Food Quality and Preference* 28 (2), 492–504.

- 603 Olsen, G. D., Swait, J., 1997. Nothing is Important. Working Paper, September 1997, Advanis Inc.,
604 12 W University Ave. #205, Gainesville, FL 32601.
- 605 Ortega, D. L., Wang, H. H., Wu, L., Olynk, N. J., 2011. Modelling heterogeneity in consumer
606 preferences for select food safety attributes in China. *Food Policy* 36 (2) 318–324.
- 607 Rappoport, L., Peters, G. R., Downey, R., McCann, T., Huff-Corzine, L., 1993. Gender and age
608 differences in food cognition. *Appetite* 20 (1), 33–52.
- 609 Richards, T. J., Padilla, L., 2009. Promotion and fast food demand. *American Journal of Agricul-*
610 *tural Economics* 91 (1), 168–183.
- 611 Rigby, D., Balcombe, K. Burton, M., 2009. Mixed logit model performance and distributional
612 assumptions: preferences and GM foods. *Environmental and Resource Economics* 42 (3) 279–
613 295.
- 614 Rose, J. M., Bliemer, M. C., 2009. Constructing efficient stated choice experimental designs. *Trans-*
615 *port Reviews* 29 (5), 587–617.
- 616 Train, K., 2009. *Discrete choice methods with simulation*. Cambridge University Press.
- 617 Vij, A. and Walker, J.L. (forthcoming). Hybrid choice models: The identification problem. *The*
618 *Handbook of Choice Modelling*, eds. S. Hess and A. Daly.
- 619 Wansink, B., Sobal, J., 2007. Mindless eating: the 200 daily food decisions we overlook. *Environ-*
620 *ment and Behavior* 39 (1), 106–123.
- 621 Yáñez, M. F., Raveau, S., Ortúzar, J. de Dios., 2010. Inclusion of latent variables in mixed logit
622 models: modelling and forecasting. *Transportation Research Part A: Policy and Practice* 44 (9),
623 744–753.