

Response latency in stated choice experiments: impact on preference, variance and processing heterogeneity

DANNY CAMPBELL, MORTEN RAUN MØRKBAK AND SØREN BØYE OLSEN

In this paper we utilise paradata relating to the response latency as a measure of the cognitive effort invested by respondents in self-administered online stated preference surveys. While the effects of response latency have been previously explored, this paper proposes a different approach. Specifically, we attempt to disentangle preference, variance and processing heterogeneity and explore whether response latency helps to explain these three types of heterogeneity. To test our methodology we use stated choice data collected via an online survey to establish consumers' preferences for various food attributes. Results from our analysis reinforce that response latency has a bearing on the estimates of error variance and the utility coefficients. Our findings raise concerns about the appropriateness of assuming deterministic choice sets, and we also show a link between response latency and the consideration sets that were actually used by respondents. We further observe that the manner in which response latency is accommodated has implications for willingness to pay estimation. Importantly, results in this paper draw attention for the need to better explain the variations that exist among respondents, in terms of preferences, error variance and processing strategies, and that only focusing on one (or two) of these is likely to lead to erroneous inferences.

Key words: choice experiments, willingness to pay, food choice, online surveys, paradata, response latency, scale-adjusted latent class, independent availability logit.

JEL codes: C25, Q13.

With the surge in computing technology and the emergence and increasing availability of the internet up through the 1990s, online surveys have become an established mode of collecting stated preferences. This increasing popularity also stems from the fact that they have a number of advantages over more traditional survey modes, such as mailout paper-and-pen questionnaires, personal interviews and telephone interviews. Advantages typically mentioned in the resource economics literature (cf. [Lindhjem and Navrud 2011a,b](#); [Fleming and Bowden 2009](#); [Olsen 2009](#)) are reduced costs, increased speed of data collection, less item non-responses, ability to adjust questionnaires according to respondent answers on-the-fly, potential for broader stimuli in terms of graphics and sound, and avoidance of manual data entry mistakes. While advantages are many, this literature has also highlighted a few important disadvantages, which raise concerns regarding data quality and their suitability in non-market valuation ([Lindhjem and Navrud 2011a,b](#)). In particular, these disadvantages relate to problems concerning sample coverage and representativeness, self-selection bias, and a so-called “pure survey mode effect” (i.e., where a respondent provides different answers to otherwise identical questions only because it is administered through different survey modes).

This paper focuses on one aspect of online surveys—the length of time respondents take to complete the choice experiment (i.e., response latency). The concern is that, notwithstanding the fact, as pointed out by [Cook et al. \(2011\)](#), that online surveys have the flexibility to allow respondents “time to think” and reflect, an interviewer is not present to pace the respondent. As a result, there may be a tendency for some respondents not to exert the level of cognitive effort required to answer the questions in any meaningful way. While this concern also applies to other self-administered methods of data collection, understanding the role of response latency in online surveys is especially important because of the incentives that respondents often obtain for their continued participation in such surveys. Furthermore, as respondents

Danny Campbell is a lecturer in the Gibson Institute for Land, Food and Environment, School of Biological Sciences, Queen's University Belfast, Northern Ireland, Morten Raun Mørkbak is an associate professor in the Department of Business and Economics, University of Southern Denmark, Denmark, and Søren Bøye Olsen is an assistant professor in the Institute of Food and Resource Economics, University of Copenhagen, Denmark.

within pre-recruited online panels gain experience, the tendency to answer quickly may actually increase (Malhotra 2008). Consequently, with online surveys we may in fact increase the risks of panel attrition effects and surveying experienced respondents—whose primary motivation for participating stems from the reward they receive—who answer so quickly that their choices do not reflect their actual preferences. If this is indeed the case, this has obvious implications as it calls into question the validity of any inferences that are derived from the observed choices. Therefore, we should be particularly suspicious of short latencies in data from online surveys in which participants are motivated by an incentive for completing the survey (Bonsall and Lythgoe 2009). In spite of these issues, this subject has yet to receive much attention, which gives rise to the present study.

In an attempt to tackle this issue and to gauge the cognitive effort invested by respondents, we exploit an advantage of online surveying that is often overlooked, namely the potential to collect and utilise numerous paradata (i.e., data about the process by which the survey data was collected).¹ In particular, to explore the impact of what Schwappach and Strasmann (2006) and Olsen (2009) describe as “quick-and-dirty” responses, we use paradata relating to the choice experiment response latency. As far back as the year 1500, Desiderius Erasmus, in his collection of ancient Greek and Latin proverbs, recorded the following proverb: *tempus omnia revelat*—time reveals all things. With this in mind, it is not surprising that the impact of response time on perceptual and cognitive processes in decision-making has received considerable attention in experimental psychology, consumer research and marketing research (Haaïjer, Kamakura, and Wedel 2000; Luce 1986; Rubinstein 2007). It is, however, somewhat surprising that the potential impacts of response latency in choice experiments has only been subject to relatively few investigations (e.g., see Holmes et al. 1998; Haaïjer, Kamakura, and Wedel 2000; Rose and Black 2006; Otter, Allenby, and van Zandt 2008; Brown et al. 2008; Bonsall and Lythgoe 2009; Vista, Rosenberger, and Collins 2009; Hess and Stathopoulos 2011).

Though interest in this topic has clearly increased recently, Bonsall and Lythgoe (2009) note that there is considerable scope for more research. Our paper is intended to contribute to this area. Unlike the papers mentioned above, which have established that response latency has a significant bearing on the estimates of utility coefficients, error variance, model fit and predictions, we are interested in identifying the link between the length of time respondents required to answer the choice experiment and their preferences, variances and processing strategies (specifically, choice set generation). While establishing the link between response latency and any one of these types of heterogeneity is relatively straightforward, tackling all three simultaneously poses a challenge. Nevertheless, when only one type of heterogeneity is accounted for, there is a potential risk that the actual heterogeneity among respondents is only partially explained and, in fact, may actually be an artefact of another (unmodelled) type. For these reasons, attempts to accommodate more than one type of heterogeneity would seem justified. In this paper we use latent class modelling to separately identify the different types of heterogeneity within the sample of respondents and to explore whether response latency is a predictor of these segments. This represents a step forward in the analysis of heterogeneity, as it is the first attempt at disentangling three different types of heterogeneity and is the first paper to explore the link between response latency and processing strategies.

To test our approach we use a stated choice experiment dataset that was collected via an online survey. This was administered to a pre-recruited panel of Danish consumers and had the aim of establishing their willingness to pay (WTP) for different product attributes of honey. Results from our analysis provides further evidence that preferences and the variance of the observed factors are sensitive to response latency. We find that marginal WTP estimates are reduced at higher response latencies and, as expected, that error variance generally decreases with increasing response latency. Importantly, we shed light on the fact that quick responses are likely to be a consequence of the processing strategies adopted by respondents, since respondents who answered quickest are most likely to have only considered a subset of alternatives when they made their choices. Our analysis highlights the relevance of accommodating heterogeneity for estimation outcomes as goodness-of-fit measures are significantly improved. We find that accommodating the heterogeneity in preferences, followed by processing, has the largest impact on model fit. Nevertheless, our analysis provides strong evidence for the need to address the three types

¹While the relatively easy access to paradata (such as time stamps) also applies to computer-assisted personal interviewing and computer-assisted telephone interviewing survey modes, using online surveys can facilitate the collection of additional paradata (such as keystrokes and mouse clicks).

of heterogeneity simultaneously, both in terms of model fit and predictions of preferences, variance and processing strategies.

The remainder of the paper is structured as follows. In the next two sections, the modelling approach to investigate the role of response latency on preferences, variance and processing strategies is developed, and the outline of our empirical case-study are provided. In the subsequent sections we present, results from the analysis and a general discussion, and an overall conclusion.

Modelling approach

Starting with the conventional specification of utility, where respondents are indexed by n , chosen alternatives by i , choice occasions by t and the attributes by x respectively, we have:

$$(1) \quad U_{nit} = \beta x_{nit} + \varepsilon_{nit},$$

where β are parameters to be estimated for the attributes, ε is an *iid* type I extreme value (EV1) distributed error term, with variance $\pi^2/6\lambda^2$, and where λ is a scale parameter. Given these assumptions, the probability of the sequence of choices made by individual n can be represented by the multinomial logit (MNL) model:

$$(2) \quad \Pr(y_n) = \prod_{t=1}^{T_n} \frac{\exp(\lambda \beta x_{nit})}{\sum_{j=1}^J \exp(\lambda \beta x_{njt})},$$

where y_n gives the sequence of choices over the T_n choice occasions for respondent n , i.e., $y_n = \langle i_{n1}, i_{n2}, \dots, i_{nT_n} \rangle$.

The major advantage of the specification outlined in [equation 2](#) is its simple form for choice probabilities. However, it is based on the notion that the taste intensities for a given attribute are the same for all respondents. This is clearly a limitation. In fact, it is now widely acknowledged that models relying on this strict assumption tend to be inferior to those that facilitate this heterogeneity in preferences. In addition to this, for identification purposes, the value of λ is generally set to unity and, thus, drops out of the probability calculation. But, in cases where it is believed that there is heterogeneity in the variance of the unobserved factors among respondents, this is obviously inappropriate. The estimation of separate scale parameters may, therefore, also be warranted. The above model also assumes deterministic choice sets, meaning that it is expected that all consumers consider all options presented to them. Given the increasing evidence that respondents adopt processing strategies and other simplifying heuristics (such as considering only a subset of the available options (cf. [Campbell, Aravena, and Hutchinson 2011](#))), this may be a further unrealistic assumption. Accommodating this type of heterogeneity is also likely to be beneficial as it should help ensure that the choice models are not biased by aspects that had no bearing on respondents' choices.

For these reasons, we move to more flexible specifications. While heterogeneity has become the topic of investigation in a large proportion of recent discrete choice applications, the majority of attention has been given to exploring only one type of heterogeneity at a time. Notwithstanding the contributions that these studies have made, when only one type of heterogeneity is accounted for, there may be a potential risk that the actual heterogeneity among respondents is only partially explained. Moreover, there is also a real concern of confounding, in the sense that the type of heterogeneity that has been modelled may have been actually an artefact of another (unmodelled) type. For these reasons, attempts to accommodate more than one type of heterogeneity would seem justified. Despite this, there are few examples where more than one type of heterogeneity has been addressed. In this paper, we contribute to this literature and provide the first study focusing on the issue of exploring the three types of heterogeneity concurrently. As a further advancement, we also explore this in the context of response latency. Our rationale for this is the fact that short response latencies may reflect random decision-making and/or the adoption of simplifying heuristics, whereas high response latencies may give an indication of more informed decision-making. The argument here is that this is likely to be exhibited in the preference structures, error variance and/or processing strategies. This motivates the present study on how to appropriately identify and accommodate these issues. In this paper, we use the latent class modelling approach, which we outline below.

Preference heterogeneity

In this paper we are interested in explaining the heterogeneous nature of preferences for attributes among the sample of respondents. Such (unobserved) preference heterogeneity can be accommodated by assuming random distributions. Rather than a continuous random distribution, we opt for a finite one. The advantage of this non-parametric approach is that commonly used continuous distributions may be unsuitable for representing the distribution of preferences. Finite distributions—instead—can provide greater flexibility and have practical appeal as the results can have more intuitive meaning than the parameter and moments of the distributions retrieved from continuous parametric distributions.² We specify such a latent class model, as follows:

$$(3a) \quad \Pr(y_n) = \sum_{q=1}^Q \pi_q \prod_{t=1}^{T_n} \frac{\exp(\lambda \beta_q x_{nit})}{\sum_{j=1}^J \exp(\lambda \beta_q x_{njt})},$$

where it is assumed that respondents can be identified as belonging to a specific latent class, q , each of which differs with respect to the β parameters, hence, denoted by β_q , and where λ is, again, constrained to unity for identification purposes. The unconditional probability of belonging to class q is given by π_q . Given our interest in response latency, we specify a MNL in which membership is regressed on the combined time taken to complete all choice tasks:

$$(3b) \quad \pi_q = \frac{\exp(\omega_q + \psi_q \mathcal{L}_n)}{\sum_{q=1}^Q \exp(\omega_q + \psi_q \mathcal{L}_n)},$$

where ω_q denotes the constant corresponding to latent class q , ψ_q is the effect of response latency ($\mathcal{L}$) on membership to class q , and, where ω_{q_0} and ψ_{q_0} are constrained to be zero for identification purposes.³ The attraction of this specification is that, in addition to identifying latent classes of respondents based on their preferences, we can assess the influence of response time on membership to the latent segments.

Variance heterogeneity

Despite the attraction of assessing the influence of response time on preferences, choices made very quickly may have a higher variance compared to those that were deliberated over a longer period, hence the potential label “quick-and-dirty” for the faster responses. In this paper, we explore a specification where a distribution in the scale parameter is facilitated. We recognise that the length of time required by respondents to make a well-balanced choice varies across individuals. For this reason, there are likely to be instances where the variances of the unobserved factors are not associated with response latency. Therefore, a probabilistic approach for identifying heterogeneity in variances would seem justified. This is likely to offer a more flexible solution as the differences in variance for a specific response latency are associated with a probability.

In an attempt to uncover and explain this heterogeneity in variances, we again make use of the latent class modelling framework. Specifically, we implement a variant of the scale-adjusted latent class modelling approach outlined in [Magidson and Vermunt \(2008\)](#) and [Campbell, Hensher, and Scarpa \(2011\)](#), whereby each latent class is described by a class-specific representation of scale:

$$(4a) \quad \Pr(y_n) = \sum_{r=1}^R \pi_r \prod_{t=1}^{T_n} \frac{\exp(\lambda_r \beta x_{nit})}{\sum_{j=1}^J \exp(\lambda_r \beta x_{njt})},$$

²We do note, however, that a continuous representation could have been used. However, we favour the appeal of finite distribution and highlight that this paper is intended to be illustrative, but suggest that this is potentially an interesting extension to this modelling approach.

³We note that we use paradata relating to the response latency of the panel of choice tasks. Of course, the latency associated with each choice task could instead be used, but in our latent class models we are interested in explaining class membership at the panel (i.e., individual) level rather than cross section (i.e., observation) level. We also note that the overall response latency averages out idiosyncrasies unique to each task and is, arguably, a better construct of overall attention (cf. [Malhotra 2008](#)). Since the time respondents spend on making their choices generally drops as they progress through the experiment (cf. [Haaaijer, Kamakura, and Wedel 2000](#); [Rose and Black 2006](#)), this also helps to disentangle the issue from the potential effects of learning and fatigue.

where it is now assumed that respondents belong to different latent classes that differ with respect to the λ parameters. To facilitate more straightforward comparison, we represent λ_r using the following notation:

$$(4b) \quad \lambda_r = 1 + \eta_r,$$

where η_r denotes the difference in the scale parameter associated with class r , subject to the constraint $\eta_r > -1$. We note that, for identification purposes, we set η_{r_1} to zero. Under this specification, a negative value of η_r implies that the variance associated with class r is larger, whereas a positive value implies a lower variance. Since class membership is latent, the unconditional probability of membership associated with class r , is given by:

$$(4c) \quad \pi_r = \frac{\exp(\omega_r + \psi_r \mathcal{L}_n)}{\sum_{r=1}^R \exp(\omega_r + \psi_r \mathcal{L}_n)},$$

where again, ω and ψ are class constants and response latency covariates respectively and are subject to the same constraint that ω_{r_R} and ψ_{r_R} are zero, meaning that probabilistic estimates of the differences in variances can be uncovered for any response latency.

Processing heterogeneity

While it is possible that respondents who answered relatively quickly processed all of the information in the choice tasks, it is also conceivable that at least some of them adopted some form of decision-making heuristic. Failing to account for this is likely to be suboptimal, and perhaps lead to misguided inferences, as the model does not reflect actual choice behaviour. For this reason we extend our investigation to incorporate the heterogeneity in the processing strategies adopted by respondents. We recognise that there are many types of processing strategies, attribute non-attendance and attribute aggregation where they share a common metric (e.g., see [Hensher 2010](#), for an overview), however, we choose to focus on the processing of alternatives. Specifically, rather than rely on the assumption that respondents considered all alternatives, we acknowledge that they may have considered only a subset of alternatives.

Following ([Manski 1977](#)), a probabilistic model can be formulated to model this type of behaviour to help distinguish between the deterministic choice set, as generated by the experimental design, and the respondent's actual consideration set. For this type of analysis we extend the independent availability logit (IAL) (cf. [Swait and Ben-Akiva 1987](#); [Swait 2001](#); [Frejinger, Bierlaire, and Ben-Akiva 2009](#); [Kaplan, Shiftan, and Bekhor 2012](#); [Richardson 1982](#); [Ben-Akiva and Boccara 1995](#); [Chang, Lusk, and Norwood 2009](#), for examples). The probability of choice in the IAL model is given by:

$$(5a) \quad \Pr(y_n) = \sum_{s=1}^S \pi_s \prod_{t=1}^{T_n} \Pr(y_n | C_s),$$

where $\Pr(y_n | C_s)$ is the conditional probability of the sequence of choices given the choice set is $C_s \subseteq S$, S is the set of subsets, π_s is the probability that C_s is the 'true' choice set. Since a respondent's true consideration set cannot be known with certainty, this model assumes that choice sets are latent, and the conditional choice model is MNL:

$$(5b) \quad \Pr(y_n | C_s) = \frac{\exp(\lambda \beta x_{nit})}{\sum_{j \in C_s} \exp(\lambda \beta x_{njt})},$$

We note that the size of S grows exponentially as a function of the number of alternatives (e.g., for a universal set with J alternatives, 2^J possible choice sets need to be taken into account (including the situation where none of the alternatives were taken into account, as would be the case under random decision-making)). As noted above, the alternatives taken into account by a respondent cannot be known with certainty. However, their observed choice behaviour helps make probabilistic statements about the likelihood of competing consideration sets being their true choice set. Moreover, since respondents who answered very quickly are unlikely to have attended to all alternatives when making their choice, response

latency is also likely to help identify the consideration sets. For this reason, we again include response latency as a covariate in the class membership expression:

$$(5c) \quad \pi_s = \frac{\exp(\omega_s + \psi_s \mathcal{L}_n)}{\sum_{s=1}^S \exp(\omega_s + \psi_s \mathcal{L}_n)}.$$

We note that the IAL model outlined here has two important differences to the IAL specification used in previous studies. Firstly, our model facilitates the panel nature of the dataset. Notwithstanding the effects of learning and/or fatigue, this is arguably a more appropriate specification, since the consideration set is likely to be stable over the repeated choices. To facilitate model tractability, the IAL model, as specified in [Louviere, Hensher, and Swait \(2000\)](#), is based on the restrictive assumption that the presence/absence of one alternative in the choice set is independent of the presence/absence of another alternative. Our second alteration to the IAL model is the derivation of probabilistic estimates for each of the possible choice set structures directly. It is anticipated that this should be better suited for retrieving estimates of the subsets of alternatives respondents actually considered.

Accounting for more than one type of heterogeneity

The above specifications provide a first step at looking at the three types of heterogeneity as well as the role that response latency plays. However, each assumes that only one type of heterogeneity is at play. This may be considered as a somewhat stringent assumption, since it is conceivable that the variations across respondents are not confined to one type. Despite this, the majority of discrete choice analysis addresses only one aspect of heterogeneity, and few studies explore two concurrently. To the best of our knowledge, no paper has yet to fully address all three simultaneously. In this paper we attempt to tackle this. To do this we expand the previous latent class models, as follows:

$$(6a) \quad \Pr(y_n) = \sum_{z=1}^Z \pi_z \prod_{t=1}^{T_n} \frac{\exp(\lambda_z \beta_z x_{nit})}{\sum_{j \in C_s} \exp(\lambda_z \beta_z x_{njt})},$$

where each of the z classes now describes a particular structure of preferences, variances and processing strategy. Similar to the above representations, unconditional class probabilities, which accommodate the role of response latency, are obtained using:

$$(6b) \quad \pi_z = \frac{\exp(\omega_z + \psi_z \mathcal{L}_n)}{\sum_{z=1}^Z \exp(\omega_z + \psi_z \mathcal{L}_n)}.$$

We note that to separately identify preference, variance and processing heterogeneity, it is necessary to set $Z = Q \times R \times S$ and, importantly, to set equality constraints (cf. [Scarpa et al. 2009](#), for further details) on the class parameters. As an example with $Q = 2$, $R = 2$ and $S = 2$, we would have $Z = 8$, and the parameters within each class would be restricted as follows:

$$(6c) \quad Z = \begin{cases} \text{class } z_1 \text{ relates to the case } \beta_{q_1}, \lambda_{r_1} \text{ and } C_{s_1}; \\ \text{class } z_2 \text{ relates to the case } \beta_{q_1}, \lambda_{r_1} \text{ and } C_{s_2}; \\ \text{class } z_3 \text{ relates to the case } \beta_{q_1}, \lambda_{r_2} \text{ and } C_{s_1}; \\ \text{class } z_4 \text{ relates to the case } \beta_{q_1}, \lambda_{r_2} \text{ and } C_{s_2}; \\ \text{class } z_5 \text{ relates to the case } \beta_{q_2}, \lambda_{r_1} \text{ and } C_{s_1}; \\ \text{class } z_6 \text{ relates to the case } \beta_{q_2}, \lambda_{r_1} \text{ and } C_{s_2}; \\ \text{class } z_7 \text{ relates to the case } \beta_{q_2}, \lambda_{r_2} \text{ and } C_{s_1}; \\ \text{class } z_8 \text{ relates to the case } \beta_{q_2}, \lambda_{r_2} \text{ and } C_{s_2}, \end{cases}$$

where, λ_{r_1} is, again, set to unity for identification purposes. The model outlined in [equation 6](#) is a fully general model for examining heterogeneity—no heterogeneity is accommodated, as in the MNL model when Q , R and S are equal to 1; one type can be uncovered in isolation when either Q , R or S is greater

than 1; two types can be retrieved when either Q , R or S is equal to 1 and the remaining two are greater than 1; and, finally, the case where all three types of heterogeneity are tackled simultaneously is when Q , R and S are all greater than 1. Therefore, given the correct equality constraints across specific classes it is possible to come up with eight model forms, depending on the assumptions of heterogeneity in respondent's preferences, or tastes (T), variances (V) and processing strategies (P):

- MNL:** A standard MNL model, where $Q = 1$, $R = 1$ and $S = 1$;
- LC(T):** A latent class model for preference heterogeneity, where $Q > 1$, $R = 1$ and $S = 1$;
- LC(V):** A latent class model for variance heterogeneity, where $Q = 1$, $R > 1$ and $S = 1$;
- LC(P):** A latent class model for processing heterogeneity, where $Q = 1$, $R = 1$ and $S > 1$;
- LC(TV):** A latent class model for preference and variance heterogeneity, where $Q > 1$, $R > 1$ and $S = 1$;
- LC(TP):** A latent class model for preference and processing heterogeneity, where $Q > 1$, $R = 1$ and $S > 1$;
- LC(VP):** A latent class model for variance and processing heterogeneity, where $Q = 1$, $R > 1$ and $S > 1$;
- LC(TVP):** A latent class model for preference, variance and processing heterogeneity, where $Q > 1$, $R > 1$ and $S > 1$.

In this paper we test each of these models against an empirical dataset. For each model, we compare model performance (in terms of model fit) as well as model output (in terms of predicted distributions of preferences, willingness to pay, variances of the unobserved factors and processing strategies). The empirical dataset that we use is outlined in the following section.

Case-study: willingness to pay for attributes of honey

Data for the present study were gathered using an online stated choice experiments focusing on Danish consumers' preferences and willingness to pay for honey. The method was found to be particularly suitable because of its ability to uncover the relative weighting of the various characteristics of honey. The attributes of honey deemed to be of greatest relevance to Danish consumers were identified as 'origin', 'method of production' and 'type of honey'. Furthermore, a cost attribute, with eight different levels, was included to denote the price increase over standard honey. An overview of all attributes and levels are presented in [table 1](#).

Using a D -efficiency criterion for evaluation, a Bayesian updated experimental design was employed using priors from a pilot study with 104 respondents. The final experimental design consisted of 12 choice tasks, each of which comprised of two generic alternatives, labelled Option A and Option B respectively, and a baseline alternative, denoted by Status-quo, the later representing a base 450 g jar of honey defined as conventional mix honey produced outside Europe and pricing DKK 25.

Table 1: Attributes and levels used in the choice experiment

Attributes	Levels (coding)
Price increase per 450 g jar (DKK)	0, 3, 8, 15, 23, 30, 40 and 55 (Price)
Origin	Local
	Denmark } Produced in Denmark (Danish)
	European
	Outside Europe } Not produced in Denmark (Non_Danish)
Method of production	Organically produced (Organic)
	Not organically produced (Non_organic)
Type of honey	Heather (Heather)
	Clover
	Rape } Non-heather (Non_heather)
	Mixture }

Respondents were sampled from a pre-recruited internet panel between June and August 2010. The sample on which the analysis in this paper is based consists of a total of 592 respondents, thus leading to 7,104 observations for model estimation. The socio-demographic distribution of the sample was approximately in accordance with that of the Danish population.

Results and discussion

Estimation results

Initial estimation results

Results from the basic MNL model (equation 2) are presented in table 2. An examination of these show that all attribute coefficients are statistically significant. We further note that the signs of the estimated coefficients are as expected—with positive signs for ‘Danish’, ‘Organic’, and ‘Heather’ and a negative sign associated with the ‘Price’ attribute. Exploring a little further would suggest that Danish consumers have highest preference for honey that is produced in Denmark, closely followed by organic, and with honey produced from heather somewhat lower.

Given our special interest in response latency, we explore how well this model performs across the range of response times. For this reason, in figure 1 we plot each respondent’s contribution to the MNL log-likelihood (LL)⁴ against the response latency associated with the choice experiment exercise. We further add histograms of these to provide additional insight. Looking firstly at the histogram for the contribution to the overall model LL, we see quite a wide range. The lowest value is found to be almost –20 (which represents an average predicted probability of 0.19), whereas the highest is –6.490 (representing an average predicted probability of 0.58). The mean and median are found to be –11.347 and –11.247 respectively.

Moving our attention to the histogram associated with latency, we find quite a range in response times. As noted by Bonsall and Lythgoe (2009), such a range can be expected since response latency varies between individuals, depending on their personal decision-making styles, and is likely to reflect circumstances, such as the extent of any distraction, the time pressure they are under and their current mental and motivational state. The median response latency associated with completing all 12 choice tasks is just over 4 minutes. While just over 90 percent of respondents completed the choice experiment within 8 minutes, a few spent almost 30 minutes completing the exercise (i.e., almost an average of 2.5 minutes per choice task). While this could of course be interpreted as respondents spending a huge effort on answering the choice sets, we cannot rule out that this is due to measurement error in terms of respondents not focusing only on the choice experiment during that period. As response latency is measured solely on a click-by-click basis, we have no way of telling whether the respondents faced any distractions or were multitasking when they were completing the choice experiment. At the other end

⁴Each respondent’s contribution to the MNL LL can be retrieved using $LL_n = \sum_{t_n=1}^{T_n} \ln(\Pr(i_{nt}|\hat{\beta}, x_n))$.

Table 2: Estimation results (MNL model)

	est.	t-rat.
β_{Price}	-0.040	30.05
β_{Danish}	0.876	19.63
β_{Organic}	0.827	28.28
β_{Heather}	0.353	8.28
$LL(\hat{\beta})$	-6,717.324	
K	4	
$\bar{\rho}^2$	0.139	
AIC	13,442.648	
BIC	13,447.517	

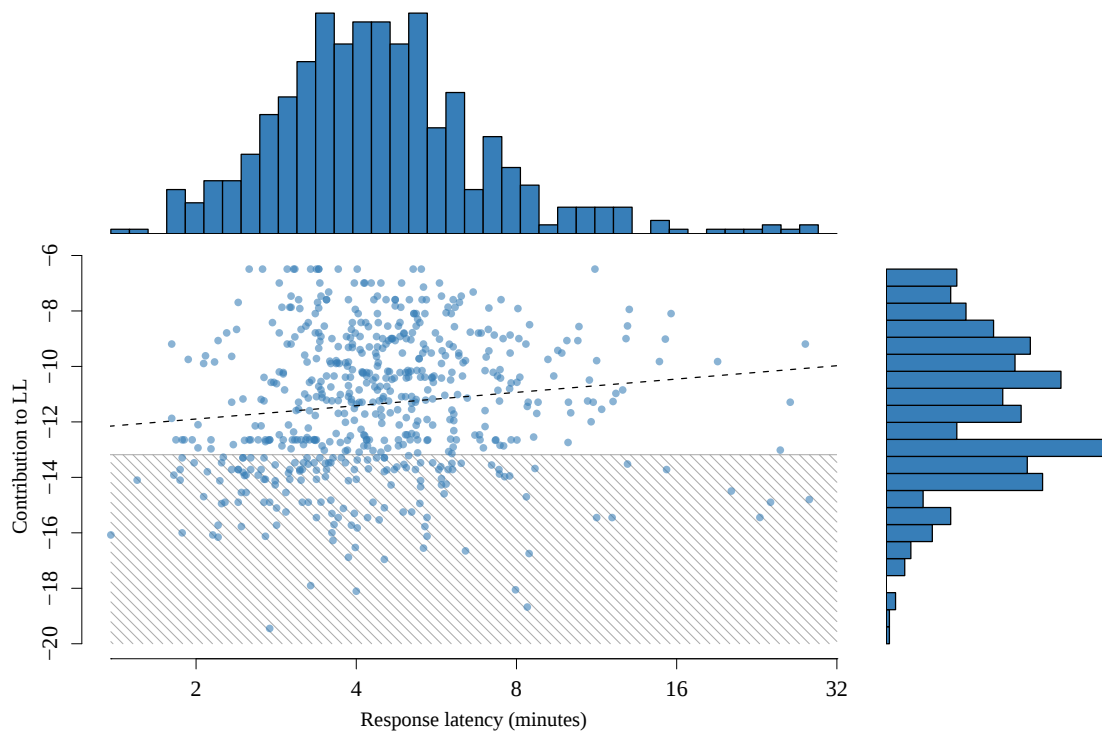


Figure 1: Response latency versus MNL contribution to log-likelihood

of the spectrum almost 20 percent of respondents completed the 12 choice tasks in less than 3 minutes (which is equivalent to an average of 15 seconds per task). While it is clearly not trivial to determine exactly what the minimum required response latency would be in order for respondents to fully evaluate all of the information contained within the choice task, response times of 15 seconds per choice task would, *prima facie*, seem to be quite short. It is also of concern that these relatively fast responses may, at best, provide nothing but noise to our survey or, at worst, bias our results.

As portrayed by the trend line (dashed line), we do see a general positive relationship (with some deviation), indicating that the MNL model better predicts the choices made by respondents who spent longer time completing the online choice experiment. We also note that the Spearman's ρ and Kendall's τ coefficients are both positive and highly significant (p -value < 0.01 in both cases), which confirms this positive correlation. This provides an indication that the MNL model predicts progressively better as the response time increases. Thus, in other words, the MNL model is less well suited for describing the choices made by respondents who answered quickly. While it may have been possible that respondents who answered relatively quickly processed all of the information in the choice tasks and made a utility maximising choice, it is also conceivable that they adopted some form of decision-making heuristic, or even made completely random choices. We note that we observe this latter possibility—as shown in the shaded region of the plot, as latency increases there is a general decrease in the number of cases where the respondent's contribution to the LL falls below the null LL associated with a random sequence of choices (i.e., $\ln(1/3) \times 12$). These findings motivate our search for more flexible models and to assess the potential of using the response latencies to help describe the heterogeneity in preferences, error variances and processing strategies.

Heterogeneity estimation results

Results from seven latent class models are presented in [table 3](#). While each model attempts at uncovering the heterogeneity among respondents, the type of heterogeneity captured differs. The first three models, labelled LC(T), LC(V) and LC(P) assume that only one type of heterogeneity is present, either relating to preferences ([equation 3](#)), variances ([equation 4](#)) or processing strategies ([equation 5](#)) respectively. The

remaining four models all assume that more than one type of heterogeneity exists among respondents (equation 6). The first three of these, identified by LC(TV), LC(TP) and LC(VP), take into account two types of heterogeneity. The final model, labelled LC(TVP), attempts at explaining all three types of heterogeneity. Central to this paper is the link between response latency and heterogeneity. After testing various specifications, we found that using the natural log of response latency as the covariate provided the best results.

In this paper we assume that the heterogeneity in respondents' preferences can be adequately explained using two latent classes. Similarly, we consider two latent classes should be sufficient to describe variance heterogeneity. For processing heterogeneity, we have omitted choice tasks that do not consist of at least two alternatives. Given the choice tasks relating to this empirical dataset consisted of three alternatives, this leads to a total of four consideration sets: (i) all three alternatives considered; (ii) only Option A and Option B considered; (iii) only Option A and Status-quo considered; and, (iv) only Option B and Status-quo considered.⁵

An examination of the two latent classes uncovered in LC(T) reveals two sub-groups of respondents with distinct preferences. The most noteworthy, and somewhat surprising, aspect is that the first class is described as placing a significantly negative value on honey originating within Denmark. The negative class membership constant implies that this is the minority class, but the coefficient associated with response latency is positive, which suggests that as response latency increases the proportion of respondents within this class also increases. As we move from the MNL model, which accounted for neither the panel nature of the dataset, preference heterogeneity nor the influence of response latency, to the LC(T) model we observe an improvement of over 1,136 log-likelihood units at the expense of fitting six additional parameters. The null embedded in the MNL model provides a likelihood ratio test statistic of 2,273.01 against the χ^2 critical value of 12.59 ($\chi^2_{6,0.05}$).

Moving to the results pertaining to LC(V), where the latent classes are denoted by differences in variance, and focusing firstly on the utility coefficients obtained for this model, reveals that they are broadly in line with those uncovered under the MNL model. The scale adjustment parameter estimated for the second class is found to be negative and significant, meaning that respondents associated with this class have higher variance relative to those in the first class, where scale is fixed to 1. The estimated values of the class membership constant and latency coefficient indicates that at the shortest response latency of 1.38 minutes (or 83 seconds) almost 80 percent of respondents are predicted as belonging in the class associated with higher variance. This is in contrast to the prediction of less than 60 percent calculated at the median latency of 4.23 minutes (or 254 seconds) and the prediction of approximately 75 percent at the highest response latency. While these findings confirms our suspicion that the very fast choices tend to introduce a lot of noise, the results also imply that a considerable proportion (i.e., 25 percent) of the choices made by respondents who took the most time were associated with relatively high variance. This is consistent with our inferences made in relation to figure 1, which showed that even for relatively high response latencies there is still a fraction of respondents for whom the model predicts worse than assuming a random sequence of choices. Finally, while we find that the LC(V) model fit is much superior to that obtained for the MNL model, it is inferior to that relating to the LC(T) model.

⁵We acknowledge that we do not conduct conventional latent class specification searches to establish the number of latent classes required to best describe the three types of heterogeneity. Nevertheless, it is sufficient for the purpose at hand of exploring the three types of heterogeneity simultaneously without further exploding the number of classes to be estimated in the final model.

Table 3: Estimation results (heterogeneity models)

	LC(T)		LC(V)		LC(P)		LC(TV)		LC(TP)		LC(VP)		LC(TVP)	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
β_{Price,q_1}	-0.093	18.97	-0.065	16.88	-0.056	34.29	-0.024	4.00	-0.107	18.83	-0.025	6.95	-0.053	5.96
β_{Danish,q_1}	-0.256	2.57	1.341	10.88	1.064	22.27	-0.044	1.69	-0.344	3.15	0.126	3.34	-0.120	2.21
$\beta_{Organic,q_1}$	1.116	15.75	1.637	18.57	0.658	21.80	0.289	4.00	1.178	15.35	0.441	6.91	0.599	6.14
$\beta_{Heather,q_1}$	0.440	5.83	0.509	6.20	0.317	7.24	0.094	3.23	0.507	6.21	0.090	3.56	0.218	4.24
β_{Price,q_2}	-0.030	14.39	-	-	-	-	-0.022	8.55	-0.047	19.87	-	-	-0.012	1.94
β_{Danish,q_2}	1.957	24.37	-	-	-	-	1.529	9.33	1.997	25.50	-	-	0.503	1.95
$\beta_{Organic,q_2}$	1.330	27.25	-	-	-	-	0.955	9.20	1.018	22.79	-	-	0.255	1.97
$\beta_{Heather,q_2}$	0.180	2.80	-	-	-	-	0.123	2.80	0.195	3.03	-	-	0.053	1.73
η_{q_1,r_2,s_1}	-	-	-0.686	31.79	-	-	4.131	3.59	-	-	2.580	5.39	1.744	4.45
η_{q_1,r_2,s_2}	-	-	-	-	-	-	-	-	-	-	<-0.999	568.18	0.931	0.01
η_{q_1,r_2,s_3}	-	-	-	-	-	-	-	-	-	-	1.483	1.20	2.136	1.71
η_{q_1,r_2,s_4}	-	-	-	-	-	-	-	-	-	-	0.907	1.47	1.109	1.67
η_{q_2,r_2,s_1}	-	-	-	-	-	-	1.173	6.15	-	-	-	-	4.611	1.72
η_{q_2,r_2,s_2}	-	-	-	-	-	-	-	-	-	-	-	-	3.181	1.54
η_{q_2,r_2,s_3}	-	-	-	-	-	-	-	-	-	-	-	-	-0.126	0.00
η_{q_2,r_2,s_4}	-	-	-	-	-	-	-	-	-	-	-	-	0.553	0.01
ω_{q_1,r_1,s_1}	-0.947	3.14	-1.529	4.00	-0.827	0.57	-3.266	4.38	10.070	6.11	-1.071	0.74	7.245	0.50
ω_{q_1,r_1,s_2}	-	-	-	-	-0.981	0.68	-	-	4.921	2.17	-4.166	2.68	4.028	0.28
ω_{q_1,r_1,s_3}	-	-	-	-	-2.565	1.10	-	-	9.303	2.76	0.669	0.16	11.783	0.80
ω_{q_1,r_1,s_4}	-	-	-	-	0.000	fixed	-	-	11.810	7.98	-10.000	2.88	-0.460	0.01
ω_{q_1,r_2,s_1}	-	-	0.000	fixed	-	-	-1.365	3.05	-	-	-2.415	1.67	8.540	0.59
ω_{q_1,r_2,s_2}	-	-	-	-	-	-	-	-	-	-	-0.589	0.41	-3.913	0.03
ω_{q_1,r_2,s_3}	-	-	-	-	-	-	-	-	-	-	-4.394	1.50	6.149	0.42
ω_{q_1,r_2,s_4}	-	-	-	-	-	-	-	-	-	-	0.000	fixed	10.667	0.73
ω_{q_2,r_1,s_1}	0.000	fixed	-	-	-	-	-2.106	3.11	10.308	6.36	-	-	6.931	0.48
ω_{q_2,r_1,s_2}	-	-	-	-	-	-	-	-	10.960	6.40	-	-	5.938	0.41
ω_{q_2,r_1,s_3}	-	-	-	-	-	-	-	-	-4.720	0.00	-	-	-3.967	0.09
ω_{q_2,r_1,s_4}	-	-	-	-	-	-	-	-	0.000	fixed	-	-	-3.010	0.02
ω_{q_2,r_2,s_1}	-	-	-	-	-	-	0.000	fixed	-	-	-	-	8.973	0.62
ω_{q_2,r_2,s_2}	-	-	-	-	-	-	-	-	-	-	-	-	9.909	0.68
ω_{q_2,r_2,s_3}	-	-	-	-	-	-	-	-	-	-	-	-	-3.935	0.03
ω_{q_2,r_2,s_4}	-	-	-	-	-	-	-	-	-	-	-	-	0.000	fixed

Continued on next page

Table 3: Estimation results (heterogeneity models) (continued from previous page)

	LC(T)		LC(V)		LC(P)		LC(TV)		LC(TP)		LC(VP)		LC(TVP)	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
ψ_{q_1,r_1,s_1}	0.568	2.84	0.787	3.00	4.154	2.73	1.513	3.09	9.592	23.03	3.472	2.33	9.058	0.60
ψ_{q_1,r_1,s_2}	-	-	-	-	3.569	2.34	-	-	10.787	12.62	5.189	3.40	10.035	0.67
ψ_{q_1,r_1,s_3}	-	-	-	-	1.895	0.87	-	-	6.937	3.72	-3.680	0.60	0.494	0.03
ψ_{q_1,r_1,s_4}	-	-	-	-	0.000	fixed	-	-	5.189	3.33	6.446	3.32	-2.901	0.05
ψ_{q_2,r_1,s_1}	-	-	0.000	fixed	-	-	1.233	3.53	-	-	5.059	3.39	8.934	0.59
ψ_{q_2,r_1,s_2}	-	-	-	-	-	-	-	-	-	-	2.895	1.94	-2.384	0.01
ψ_{q_2,r_1,s_3}	-	-	-	-	-	-	-	-	-	-	3.266	1.39	7.717	0.51
ψ_{q_2,r_1,s_4}	-	-	-	-	-	-	-	-	-	-	0.000	fixed	4.465	0.30
ψ_{q_1,r_2,s_1}	0.000	fixed	-	-	-	-	1.308	2.99	9.297	25.24	-	-	8.666	0.58
ψ_{q_1,r_2,s_2}	-	-	-	-	-	-	-	-	8.587	15.21	-	-	9.055	0.60
ψ_{q_1,r_2,s_3}	-	-	-	-	-	-	-	-	-1.985	0.00	-	-	-1.910	0.01
ψ_{q_1,r_2,s_4}	-	-	-	-	-	-	-	-	0.000	fixed	-	-	-6.415	0.06
ψ_{q_2,r_2,s_1}	-	-	-	-	-	-	0.000	fixed	-	-	-	-	8.506	0.57
ψ_{q_2,r_2,s_2}	-	-	-	-	-	-	-	-	-	-	-	-	7.625	0.51
ψ_{q_2,r_2,s_3}	-	-	-	-	-	-	-	-	-	-	-	-	-2.129	0.01
ψ_{q_2,r_2,s_4}	-	-	-	-	-	-	-	-	-	-	-	-	0.000	fixed
LL($\hat{\beta}$)	-5,580.817		-6,613.553		-6,223.224		-5,483.118		-5,375.352		-6,042.221		-5,287.780	
K	10		7		10		16		22		22		46	
$\bar{\rho}^2$	0.284		0.152		0.201		0.295		0.308		0.223		0.317	
AIC	11,181.633		13,241.105		12,466.448		10,998.235		10,794.705		12,128.441		10,667.560	
BIC	11,180.501		13,242.974		12,465.316		10,991.104		10,781.573		12,115.310		10,630.428	

The LC(P) model has utility coefficients as well as coefficients describing the availability function. Inferences relating to sign and significance of the utility coefficients would appear to be somewhat similar to those reached under the MNL model. Of greater interest are the availability function coefficients. The constants for the first three consideration sets are negative, indicating that at quick response times they have a smaller probability share compared to the final consideration set (i.e., consideration of only Option B and Status-quo). However, as indicated by the positive influence of the covariate on membership, as response latency increases the proportion of respondents predicted as having used the first three consideration sets increases. This has two important implications. Firstly, it draws attention to the fact that the assumption of deterministic choice sets is not appropriate. The second, albeit not surprising, implication is that the information respondents process is linked with the length of time they spend completing the choice experiment. Computing the probabilities of approximately 40 and 70 percent for consideration of all three alternatives at the shortest latency against the median latency respectively illustrates these points. Finally, compared to the model fit achieved under the MNL model, we find that the LC(P) model is much superior. Although the model fit is substantially lower than that associated with the LC(T) model, it is significantly higher than the LC(V) model, demonstrating that overlooking the heterogeneity in processing is of greater consequence than not accounting for the variance heterogeneity.

Our LC(TV) model attempts to accommodate heterogeneity in the preferences and error variances among the respondents. This is accomplished using a 4-class model. Classes 1–2 are estimated with the same preferences and we observe that they are akin to those retrieved for the first class in the LC(T) model, albeit of a different magnitude, which reflects the differences in scale. The positive and significant scale offset reveals that these two classes can be distinguished in terms of error variance, with the second class having significantly less variance compared to the first class. Classes 3–4 are also estimated with independent scale parameters but have equality constraints in place with respect to the utility coefficients. Again, we find comparisons can be made against the preference structure uncovered in the LC(T) model, but we remark that respondents belonging in these classes could be further segmented in terms of error variance. Inspecting the class membership coefficients implies that the fourth class is the largest at short response times, but that the other size of the classes increases with increased response times. The model fit is much improved, compared to the fits achieved under either LC(T) or LC(V). In fact, at the expense of 6 and 9 additional parameters, we find increases of 98 and 1,130 log-likelihoods against the LC(T) and (LC(V) models respectively, which contributes to a significant gain in fit, irrespective of base model. This highlights the disadvantage and inappropriateness of assuming that the only differences in respondents are in their preferences or error variances, but not in both.

The LC(TP) model assumes homoscedasticity across respondents but recognises the fact that there may be variation in their preferences and processing strategies. This is based on an 8-class model: with separate equality constraints for the utility coefficients in classes 1–4 and classes 5–8, classes 1 and 5 whereby all three alternatives were considered, classes 2 and 6 relating to the consideration set containing only Option A and Option B, classes 3 and 7 capture the case where the consideration set was limited to only Option A and Status-quo, finally, classes 4 and 8 where only Option B and Status-quo were part of the respondent's actual choice set. Looking firstly at the utility coefficient reveals that similar inferences can be made as those reached under LC(T). Again, we find evidence to support the fact that the assumption of deterministic choice sets may be misguided and that the length of time respondents spent answering the choice experiment helps predict the consideration sets actually used by respondents. We remark that the model fit is significantly higher than all previous models, which further makes evident that the repercussions on model performance of neglecting to account for this type of processing strategy are more abundant than when the heterogeneity in variances is overlooked.

The LC(VP) model recognises the fact that the error variance may be different depending on the consideration set adopted by respondents. But for reasons of confounding, it is not possible to retrieve separate scale parameters for each consideration set. To address this issue, we, again, use an 8-class specification, and where the variations across respondents are limited to error variances and processing strategies. A similar class structure is used as the previous model, but where classes 5–8 are estimated as having a different scale parameter to classes 1–4, which have all been fixed to unity to facilitate identification. While inferences regarding the utility coefficients remain largely unchanged, we see that respondents who considered all alternatives and those who considered only Option A and Option B can be further segmented on the basis of error variance. For the remaining respondents, the insignificant scale offset parameters implies homoscedasticity. The parameters relating to class membership supports the

fact that error variances and processing of information is correlated with response latency. While we observe a significant improvement in model fit relative to the LC(V) and LC(P) models, the fact that it is substantially inferior to those attained under both the LC(TV) and LV(TP) models, again, highlights the more important role of preference heterogeneity.

When exploring heterogeneity it seems logical to consider all three types, otherwise there is a concern that the observed differences could be a result of different relative preferences, different error variances, different processing strategies, or all three. The LC(TVP) aims at isolating the different types of heterogeneity across the sample of respondents, which leads to a 16-class model. Classes 1–8 are as defined in the LC(VP) model, whereas classes 9–16, while having a similar structure of variances and processing strategies, are based on separate utility coefficient. Looking firstly at the utility coefficients we find a similar pattern of preference heterogeneity emerging. Similarly, we find signs of significant heterogeneity in error variance and processing. While we remark that the large number of classes impedes our ability to make any inferences on the class composition without further analysis, we draw attention to the obvious improvement in model fit, which is now significantly higher than all other specifications. We note that this improvement is also supported by the $\bar{\rho}^2$, AIC and BIC statistics, even after accounting for the large increase in additional parameters.

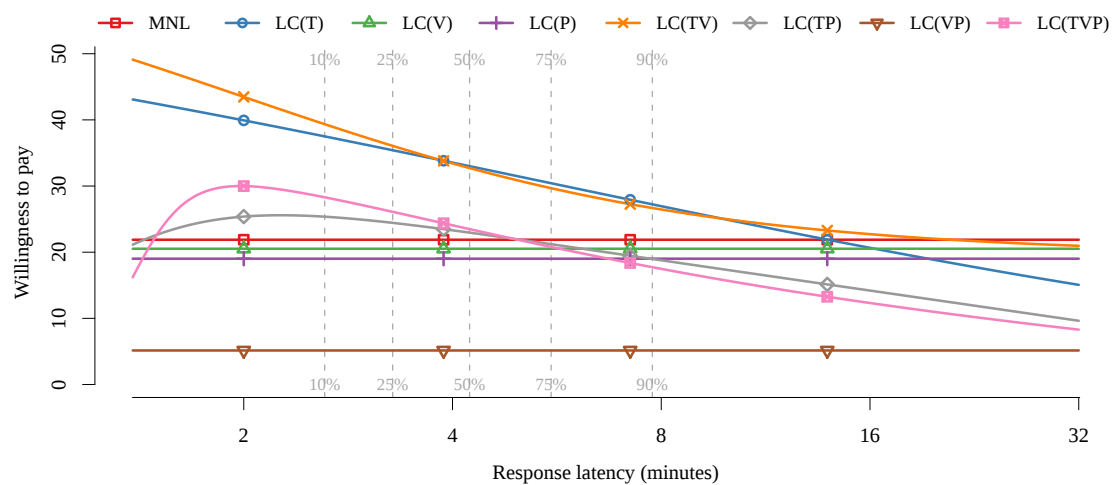
As alluded to above, interpreting the membership coefficients is not straightforward, especially as the number of classes increase. However, this is necessary if we are to get an impression of the heterogeneity and to understand the role of response latency as well as the impact of model specification. For this reason in the following sections we illustrate the heterogeneity uncovered from each of the models.

Response latency and preference heterogeneity

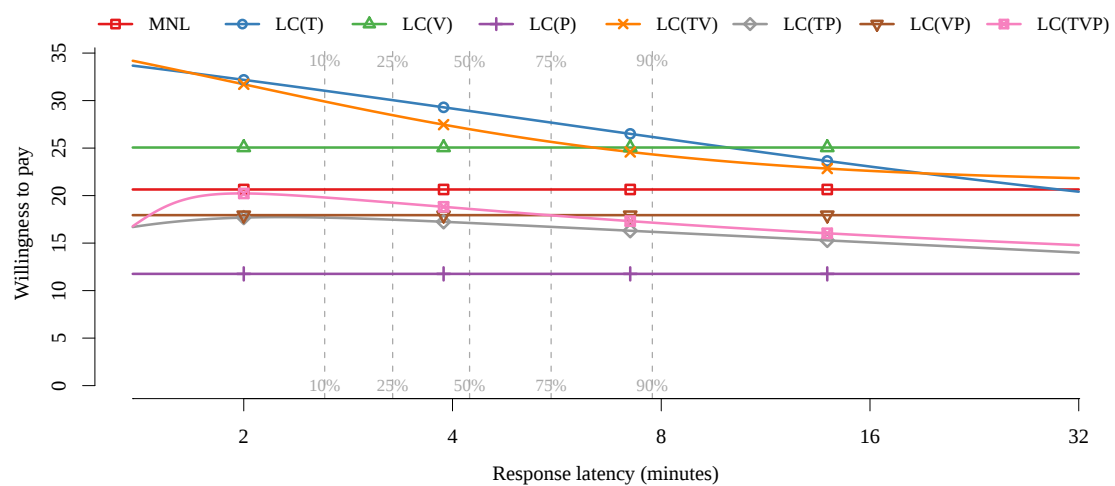
Any meaningful comparison of preference heterogeneity across the various models is not possible, since each model is subject to a different scaling. What does make comparative sense is heterogeneity in the implied marginal WTP estimates, since the scale effect is neutralised. In [figure 2](#) we plot the marginal WTP estimates for each model against response latency for the three honey attributes. We note that we have weighted the marginal WTP according to the unconditional class membership probabilities (and, thus, they represent the most likely marginal WTP at the given latency).

Given the assumption of preference homogeneity in the MNL, LC(V), LC(P) and LC(VP) models the WTP estimates remain constant irrespective of response latency. For the models that do facilitate preference heterogeneity there are clear differences in the estimates across response times (reflecting the shifts in class membership with latency). For the ‘Danish’ ([figure 2\(a\)](#)) and ‘Organic’ ([figure 2\(b\)](#)) attributes we observe a general shift downwards in marginal WTP as response latency increases—depending on specification marginal WTP estimates fall from around DKK 50 and DKK 35 to around DKK 10 and DKK 15 respectively between the shortest and longest response times. Notwithstanding the general decrease, it is interesting to note that both models that incorporate preference and processing heterogeneity (i.e., LC(TP) and LC(TVP)) suggest increasing marginal WTP up to response times of approximately 2 minutes, after which it starts to decrease. In contrast, the estimates of marginal WTP relating to ‘Heather’ ([figure 2\(c\)](#)) appear to be relatively stable (in the region of DKK 5) across all response latencies, which implies relative homogeneity in the value the respondents placed on this attribute.

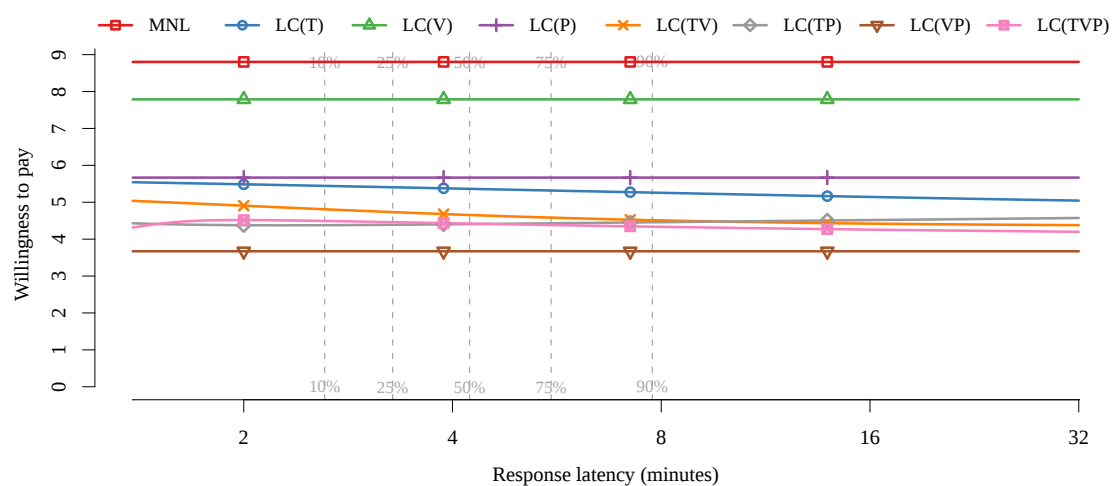
Issues of response latency aside, the estimates of marginal WTP vary according to model specification. For example, a comparison of the estimated marginal WTP at the median response latency, reveals variations between approximately DKK 5 and DKK 35 for honey originating in Denmark, a range from approximately DKK 12 to DKK 30 for organic honey, and, between around DKK 4 and DKK 9 in the case of honey produced from heather. While difficult to establish which of our models portrays the most accurate distribution of marginal WTP for the honey attributes, on the basis of model fit alone, one could argue that the LC(TVP) most accurately describes WTP. Based on this assertion, we find that aside from the LC(TP) model, none of the models seem capable of retrieving the distributions of marginal WTP. This provides further proof, at least for this empirical dataset, of the relative importance of accounting for heterogeneity in respondents’ preferences and processing strategies compared to error variance.



(a) Response latency versus marginal willingness to pay for 'Danish'



(b) Response latency versus marginal willingness to pay for 'Organic'



(c) Response latency versus marginal willingness to pay for 'Heather'

Figure 2: Response latency versus marginal willingness to pay (DKK per jar)

Response latency and variance heterogeneity

Findings presented in [table 3](#) indicate that response latency has a bearing on the variance of the unobserved factors. To illustrate this, in [figure. 3](#) we plot the predicted error variance against response latency for each of our models. For straightforward comparison, these are relative to the shortest response latency. Again, the relative variances are calculated on the basis of the unconditional class membership probabilities (and, thus, represent the most likely relative variance at the given latency).

With the exception of the MNL, LC(T), LC(P) and LC(TP) models (which assume homoscedasticity, irrespective of response latency), there appears to be a reduction in the error variance as response latency increases. Nevertheless, the rate and degree of this reduction varies across the model specifications. The most striking reduction is predicted under the LC(VP) model, which suggests that the variance associated with respondents who required 8 minutes to complete the choice experiment is less than one-fifth of that associated with those who completed the experiment in the shortest time. However, we note that this model does not account for preference heterogeneity, and may, therefore, be limited in its potential to uncover the actual reduction in variance. Interestingly, the estimates of reduction in relative variance are somewhat comparable under the LC(V) and LC(TV) models. Under these models, the variance of the unobserved factors for respondents around the median time of just over 4 minutes is approximately 75 percent of that associated with those who made their choices in the shortest time. Moving to the predictions of average reductions in error variance computed using the more flexible LC(TVP) model, we see a somewhat different pattern. While the error variance is shown to initially (and quite sharply) decrease in accordance with the previous models, it starts to increase at around 2 minutes. In fact, this increase is to the extent that respondents who required 8 minutes to complete the choice experiment had around the same error variance as those who completed very quickly. While this would appear to be a peculiar finding, it confirms our earlier suspicions that these respondents may have been distracted or were multitasking during the experiment, in which case it is reasonable to expect increases error variance. The fact that this was not picked up by the other models and that this superior fitting model differs so much from the previous predictions of variance highlights the importance of taking the three

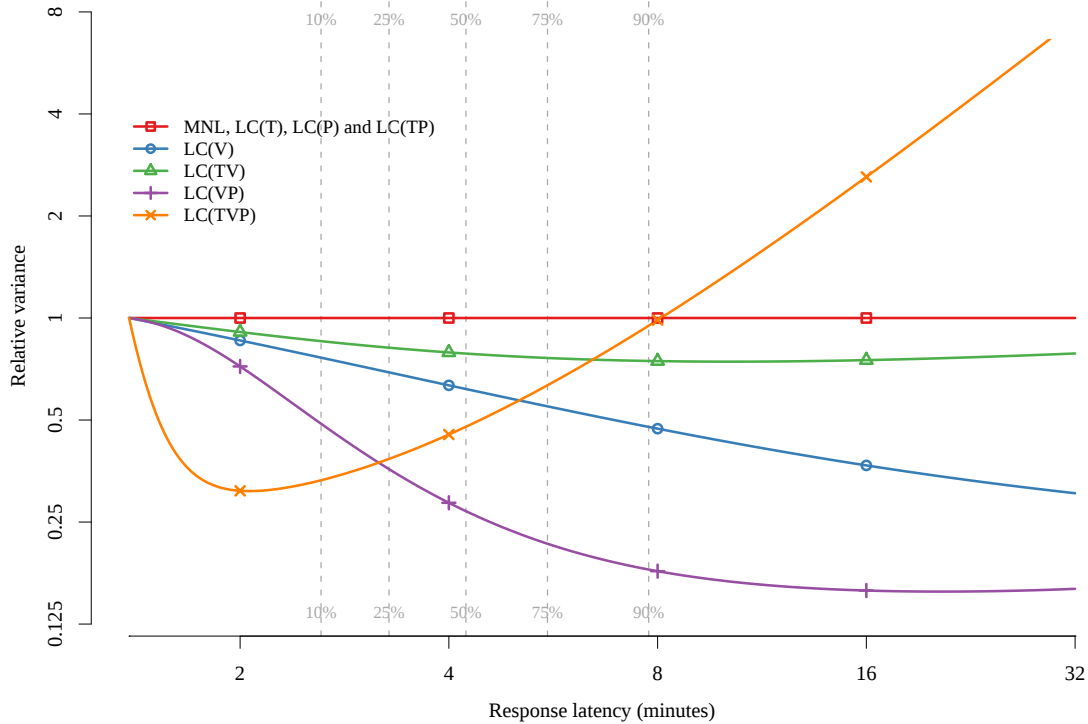


Figure 3: Response latency versus relative variance across all models

types of heterogeneity into account simultaneously, otherwise any inferences regarding response latency and error variance may be misleading.

Response latency and processing heterogeneity

Four of our models rely on the assumption of deterministic choice sets (namely, the MNL, LC(T), LC(V) and LC(TV) models) and this is considered to be the case across all response latencies. For the remaining models, we can use the class membership coefficients to derive probabilistic predictions of the four consideration sets assumed in our analysis. In [figure 4](#) we plot these predictions. As stated above, the MNL, LC(T), LC(V) and LC(TV) models imply 100 percent membership to the consideration set involving all three alternatives (as shown in [figure 4\(a\)](#)). However, inspection of the results attained from the models that where this is relaxed, indicate that this is likely to be an erroneous assumption. Predictions from these models are remarkably consistent and show that the proportion of respondents who actually considered all three alternatives is likely to be less than 75 percent. As would be expected, we do find that this proportion increases and at very short latencies, it is likely that more than half of respondents do not consider the full choice task. From the remaining consideration sets, [figure 4\(b\)](#) indicates that a high proportion of respondents made choices where they excluded the Status-quo. This is an important finding, as it could help explain some of the status-quo effect that is often associated with stated choice experiments. We observe that the proportion who adopted this heuristic is in the region of 30 percent. It is also worth pointing out that this type of behaviour appears to be more prevalent among respondents who answered very quickly, which is also to be expected. The final two consideration sets follow a somewhat similar pattern and only seem to be adopted by respondent who provided their responses in less than 2–3 minutes.

To further tease out the processing heterogeneity, for each alternative we add together the probabilities of classes in which it forms part of the consideration set. This is aimed at providing us with a impression of how attendance to each of the alternatives changes with response latency. The results from these calculations are outlined in [figure 5](#). While we find evidence that the majority of respondents included Option A and Option B in their consideration set ([figures 5\(a\)](#) and [5\(b\)](#) respectively), the same can not be said for Status-quo ([figure 5\(c\)](#)), which is compliant with our earlier observation.

Conclusions

In questionnaire surveys there is an obvious link between the effort that respondents allocate to answering the questions and the quality of the obtained survey data. In general, it might be conjectured that the larger the effort spent on answering a question, the greater the quality of the answer. However, with the upward trend in self-administered online stated preference surveys, there is potentially an increased risk that respondents do not fully engage with the survey. With many of these surveys providing incentives to pre-recruited panels of respondents, there are further concerns of panel attrition effects and that we may be surveying experienced respondents wishing to take advantage of the incentive. The fear is that these respondents are unlikely to exert the necessary cognitive effort required to answer the questions in a meaningful way. While the relationship between respondent effort and data quality is quite obvious, it is less trivial how effort is measured and what the relationship between respondent effort and data quality actually is. In this paper we use response latency as a proxy for respondent effort.

This paper proposes a novel approach to model the effects of response latency on the estimates of utility coefficients, error variance and processing strategies. While it is relatively straightforward to establish each of these in turn, this paper set out to address them simultaneously, which is considerably more challenging. Our motivation stems from a concern that accommodating only one type of heterogeneity may provide an incomplete picture and may even be distorted by the other (unmodelled) types. Our analysis is based on the latent class modelling framework and is aimed at separately identifying the heterogeneity in preferences, error variances and the processing strategies across respondents. We include response latency as a covariate in the class membership function to enable its role to be assessed.

We test our approach using an empirical dataset collected via an online survey to establish the value that Danish consumers are willing to pay for various attributes associated with honey. Methodological issues aside, our results indicate Danish consumers are willing to pay more for value-added honey,

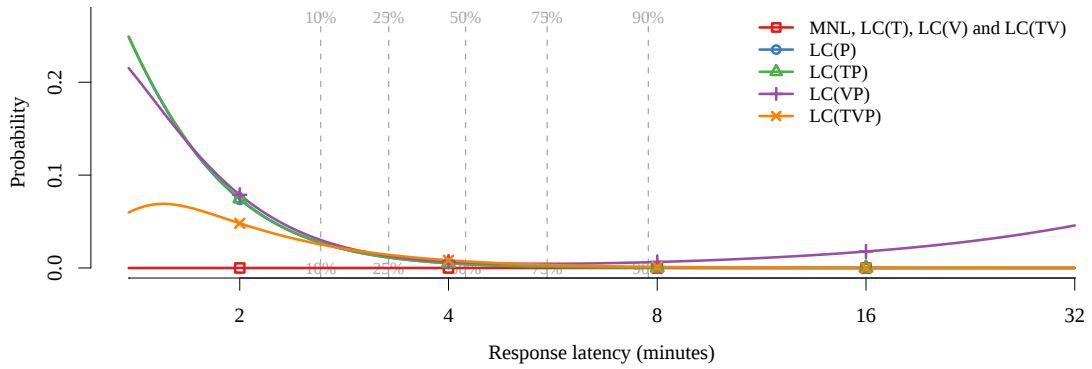
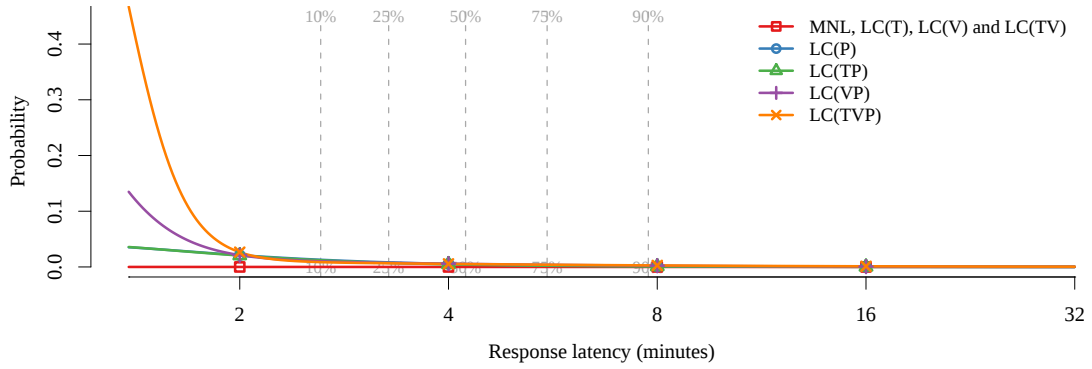
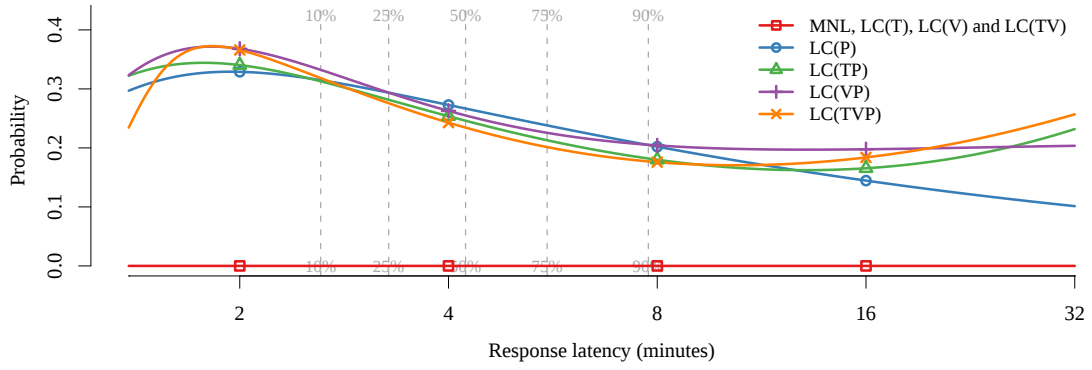
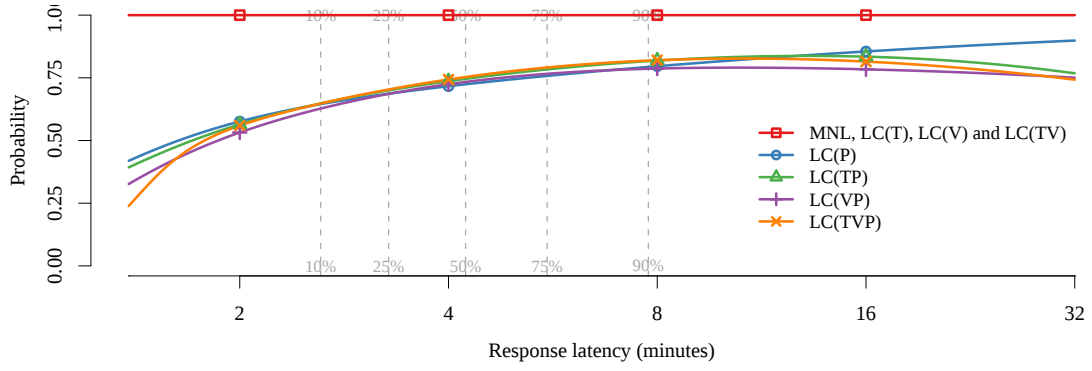
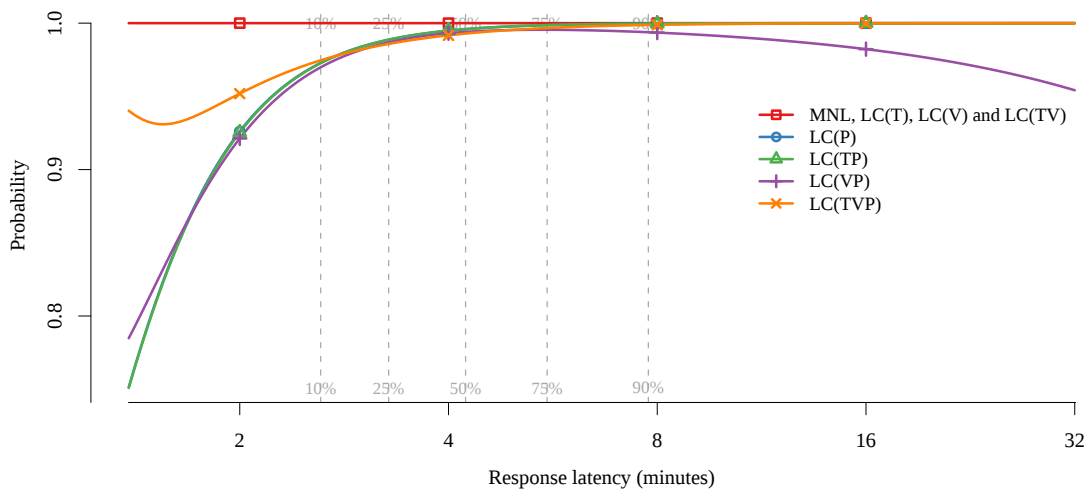
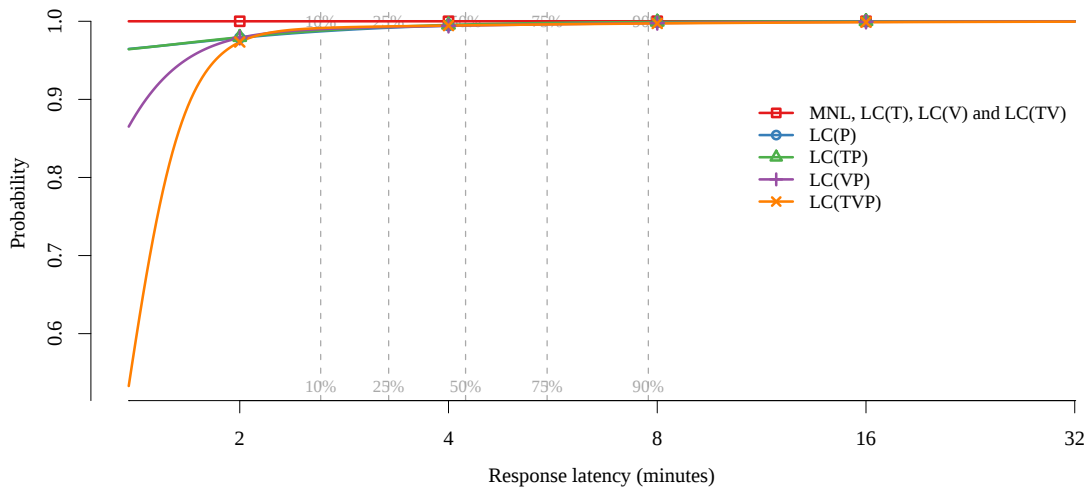


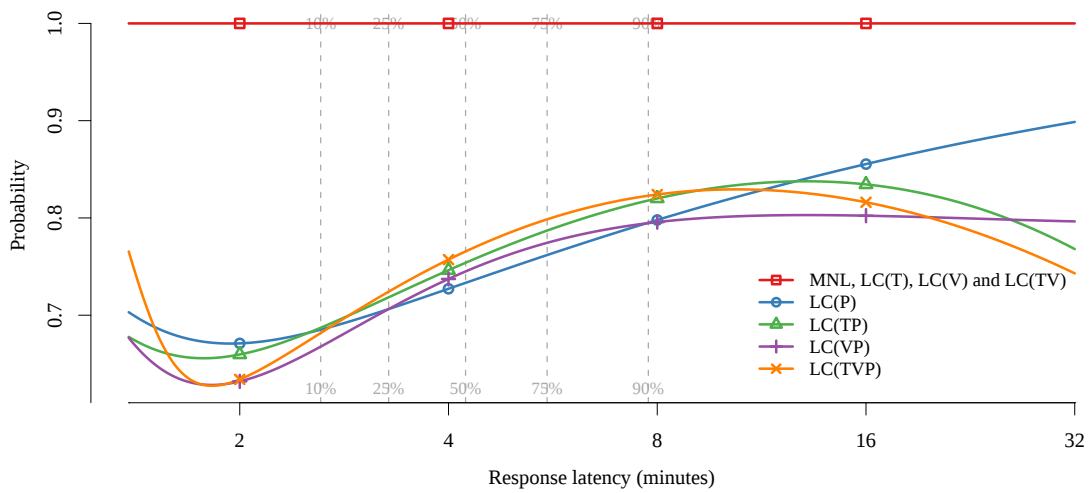
Figure 4: Response latency versus (unconditional) probabilities of alternative processing



(a) Response latency versus consideration of Option A



(b) Response latency versus consideration of Option B



(c) Response latency versus consideration of Status-quo

Figure 5: Response latency versus (unconditional) probabilities of alternative consideration

with honey originating from Denmark and honey that has been honey organically produced valued most highly.

Our paper raises a number of methodological issues. Importantly, in accordance with earlier studies (e.g., [Haaaijer, Kamakura, and Wedel 2000](#); [Rose and Black 2006](#); [Otter, Allenby, and van Zandt 2008](#); [Vista, Rosenberger, and Collins 2009](#); [Haaaijer, Kamakura, and Wedel 2000](#)) we show that response latency is an important consideration when modelling discrete choice data. We provide further evidence that preferences and the variance of the observed factors are sensitive to response latency. We find that marginal WTP estimates are reduced at higher response latencies and, as expected, that error variance generally decreases with increasing response latency. Importantly, we also shed light on the fact that quick responses are likely to be a consequence of the processing strategies adopted by respondents, since respondents who answered quickest are most likely to have only considered a subset of alternatives when they made their choices. While this seems obvious, this is the first paper implementing a model to establish the link. Our analysis highlights the relevance of accommodating heterogeneity for estimation outcomes as goodness-of-fit measures are significantly improved. We find that recognising the heterogeneity in preferences, followed by processing, has the largest impact on model fit. Nevertheless, our analysis provides strong evidence for the need to address the three types of heterogeneity simultaneously, both in terms of model fit and predictions of preferences, variance and processing strategies. This is reflected by the fact that our model which accounts for the three types of heterogeneity at once also seems better equipped to deal with the ambiguous relationship between response latency and the heterogeneity across respondents. The distributions of heterogeneity when retrieved individually appear to be misguided when compared to when they are all accounted for at once. We show that the isolation of the different types of heterogeneity significantly adds to our understanding of the variations among the sample of respondents.

From data quality and, as we have seen, from welfare analysis standpoints, it may be tempting to 'clean' datasets from respondents below (or above) a certain response latency as suggested by [Bonsall and Lythgoe \(2009\)](#). However, we stress that our analysis does not, and was not intended to, identify what these thresholds might be. Instead, our analysis is intended to provide analysts with a modelling framework to assess the sensitivity of their survey results to response latency. Moreover, our analysis highlights that defining such thresholds is a difficult judgement. We find that the link between response latency and data quality is ambiguous. Low quality data is to be found irrespective of the response latency and that, at best, it is only possible to make probabilistic statements of data quality at different response latencies.

We note that our findings are most likely not isolated to online surveys. Response latency can be ascertained in other survey modes as well. While the issue may be of a lesser consequence in other modes, due to reduced panel attrition effects, it is worth uncovering the role of response latency in different modes. This would be especially useful, given the evidence that response latency is likely to vary across modes (e.g., [Börjesson and Algiers 2011](#)), as it would provide us with additional criteria that we can evaluate when designing stated choice surveys. Even though our analysis is based on a food application, the impact of response latency and the modelling framework introduced should be of interest to a broader audience. We encourage fellow researchers to replicate our analysis to ascertain the extent to which our findings apply in other settings.

While our findings promote the further use of paradata that is easily obtainable from online surveys, more research should be directed towards how exactly the paradata is measured. In relation to response latency, as also suggested by [Lindhjem and Navrud \(2011b\)](#), there are advantages in developing more stringent measures (for instance, accounting for webpage load times as well as respondent multi-tasking would seem appropriate). With regard to the modelling framework presented here, there is also scope for further research in terms of developing even more flexible models. An obvious extension is to facilitate more latent classes. In the current analysis, for illustrative purposes, we have limited the number of classes. We acknowledge that our classes may not provide sufficient flexibility to fully accommodate the heterogeneity in preferences, error variance and processing strategies, not to mention the potential presence of attribute non-attendance or other simplifying heuristic decision rules not captured in our specification. Also, extending the model to utilise choice task latency rather than choice sequence latency may present a fruitful avenue ahead (although we do draw attention to the fact that this may come at the risk of confounding with the effects of learning and fatigue and may have implications if the panel specification is to be maintained).

Another avenue for future research is how to reduce the incidence of so called “quick-and-dirty” responses, or, in other words, how to induce respondents to spend enough time answering the choice tasks. Especially in internet based surveys it would seem that technical solutions, such as temporarily delaying the availability of the ‘next’ button and the use of popup screens warning respondents that they may be answering too fast, may provide a way of enforcing longer reflection time as suggested by Frank (2010). In a less forceful approach, encouraging respondents to take “time to think” (e.g., Cook et al. 2011; MacMillan, Hanley, and Lienhoop 2006), introduction of incentives that depend on survey engagement and respondent effort rather than merely participation, and possibly the introduction of a ‘slow talk’ script informing them that other respondents answer too quickly may also be explored. However, we warn that all these strategies may lead to unwanted side-effects, including increased drop-out rates and fatigue. Nevertheless, without further research it is difficult to establish if the advantages of these strategies outweigh their potential side-effects.

In conclusion, irrespective of response latency, our paper clearly demonstrates the need for researchers to consider the variations among respondents. Despite the significant gains that have been made in this regard, there is a need for models that are better equipped to accommodate this variation. We should recognise that heterogeneity is not restricted to preferences, nor is it limited to either variances or processing strategies. The factors influencing choice outcomes are complex, so it should not come as a surprise that accommodating as many types of heterogeneity will lead to better choice models.

References

- Ben-Akiva, M., and B. Boccara. 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing* 12(1): 9–24.
- Bonsall, P., and B. Lythgoe. 2009. Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments. *Journal of Choice Modelling* 2(2): 216–236.
- Börjesson, M., and S. Algers. 2011. Properties of internet and telephone data collection methods in a stated choice value of time study context. *Journal of Choice Modelling* 4(2): 1–19.
- Brown, T. C., D. Kingsley, G. L. Peterson, N. E. Flores, A. Clarke, and A. Birjulin. 2008. Reliability of individual valuations of public and private goods: choice consistency, response time, and preference refinement. *Journal of Public Economics* 92(7): 1595–1606.
- Campbell, D., C. D. Aravena, and W. G. Hutchinson. 2011. Cheap and expensive alternatives in stated choice experiments: are they equally considered by respondents? *Applied Economics Letters* 18(8): 743–747.
- Campbell, D., D. A. Hensher, and R. Scarpa. 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. *Journal of Environmental Planning and Management* 54(8): 1061–1076.
- Chang, J. B., J. L. Lusk, and F. B. Norwood. 2009. How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics* 91(2): 518–534.
- Cook, J., M. Jeuland, B. Maskery, and D. Whittington. 2011. Giving stated preference respondents “time to think”: results from four countries. *Environmental and Resource Economics*: 1–24.
- Fleming, C. M., and M. Bowden. 2009. Web-based surveys as an alternative to traditional mail methods. *Journal of Environmental Management* 90(1): 284–292.
- Frank, B. 2010. *The effects of enforced reflection in three simple experiments*. Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).
- Frejinger, E., M. Bierlaire, and M. Ben-Akiva. 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological* 43(10): 984–994.
- Haaijer, R., W. Kamakura, and M. Wedel. 2000. Response latencies in the analysis of conjoint choice experiments. *Journal of Marketing Research* 37(3): 376–382.
- Hensher, D. A. 2010. , *Choice Modelling The State-of-the-art and the State-of-practice - Proceedings from the Inaugural International Choice Modelling Conference*, 35–70. Emerald Press, UK.
- Hess, S., and A. Stathopoulos. 2011. Linking response quality to survey engagement: a combined random scale and latent variable approach. *University of Leeds working paper*.

- Holmes, T., K. Alger, C. Zinkhan, and E. Mercer. 1998. The effect of response time on conjoint analysis estimates of rainforest protection values. *Journal of Forest Economics* 4(1): 7–28.
- Kaplan, S., Y. Shiftan, and S. Bekhor. 2012. Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B: Methodological* 46(2): 291–304.
- Lindhjem, H., and S. Navrud. 2011a. Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics* 70(9): 1628–1637.
- Lindhjem, H., and S. Navrud. 2011b. Using internet in stated preference surveys: a review and comparison of survey modes. *International Review of Environmental and Resource Economics* 5: 309–351.
- Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. *Stated choice methods: analysis and applications*: Cambridge University Press.
- Luce, R. D. 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*, 562: New York: Oxford University Press.
- MacMillan, D., N. Hanley, and N. Lienhoop. 2006. Contingent valuation: environmental polling or preference engine? *Ecological Economics* 60(1): 299–307.
- Magidson, J., and J. K. Vermunt. 2008. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. In *Sawtooth Software Conference*.
- Malhotra, N. 2008. Completion time and response order effects in web surveys. *Public opinion quarterly* 72(5): 914.
- Manski, C. F. 1977. The structure of random utility models. *Theory and Decision* 8: 229–254.
- Olsen, S. B. 2009. Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics* 44(4): 591–610.
- Otter, T., G. M. Allenby, and T. van Zandt. 2008. An integrated model of discrete choice and response time. *Journal of Marketing Research* 45(5): 593–607.
- Richardson, A. 1982. Search models and choice set generation. *Transportation Research Part A: General* 16(5-6): 403–419.
- Rose, J., and I. Black. 2006. Means matter, but variance matter too: decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters* 17(4): 295–310.
- Rubinstein, A. 2007. Instinctive and Cognitive Reasoning: A Study of Response Times*. *The Economic Journal* 117(523): 1243–1259.
- Scarpa, R., T. J. Gilbride, D. Campbell, and D. A. Hensher. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36(2): 151–174.
- Schwappach, D. L. B., and T. J. Strassmann. 2006. “Quick and dirty numbers”? The reliability of a stated-preference technique for the measurement of preferences for resource allocation. *Journal of Health Economics* 25(3): 432–448.
- Swait, J. 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological* 35(7): 643–666.
- Swait, J., and M. Ben-Akiva. 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological* 21(2): 91–102.
- Vista, A. B., R. S. Rosenberger, and A. R. Collins. 2009. If you provide it, will they read it? Response time effects in a choice experiment. *Canadian Journal of Agricultural Economics* 57(3): 365–377.