How quick can you click? The role of response time in online stated choice experiments

Danny Campbell[a*] Morten Raun Mørkbak[b] Søren Bøye Olsen[c]

[a]Economics Division, Stirling Management School, University of Stirling, Scotland
[b]Department of Business and Economics, University of Southern Denmark
[c]Department of Food and Resource Economics, University of Copenhagen

Paper prepared for *Annual BIOECON Conference*

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 compare scale-adjusted latent class models based on preference homogeneity to those that facilitate preference heterogeneity and make further comparisons based on whether the influence of response latency is assumed to be deterministic or probabilistic. To test our methodology we use stated choice data collected via an online survey to establish German anglers' preferences for fishing site attributes in Denmark. Results from our analysis reinforce that response latency has a bearing on the estimates of error variance and the utility coefficients. Importantly, our latent class models also show that, irrespective of the length of time respondents take to complete the choice experiment, there is always a subset with high error variance (i.e., more randomness), but that this decreases as response latency increases. While estimates of willingness to pay are not affected, we observe that the manner in which response latency is accommodated has implications on the predictions of fishing trips and expected revenue.

Keywords: angling recreation; online surveys; paradata; response latency; scale-adjusted latent class; stated choice experiments
JEL classifications: C25; Q26; Q50; Q51

1 Introduction

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

*Corresponding author. Email: danny.campbell@stir.ac.uk.

in the resource economics literature (cf. Fleming and Bowden [2009], Olsen [2009] and Lindhjem and Navrud [2011a,b]) 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 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 (financial or otherwise) they receive—who answer so quickly that their choices do not reflect their actual preferences. If this is indeed the case, it 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*

[1] While 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).

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 (Luce, 1986; Haaijer, Kamakura and Wedel, 2000; 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], Haaijer, 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] and 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 propose the use of latent class models to separate the respondents who exhibit relatively random behaviour from those that invest the necessary cogitative effort. Specifically, we investigate the issue using a number of scale-adjusted latent class specifications that are intended to uncover the influence of response latency on error variance. We, subsequently, elaborate on these models to simultaneously account for the preference heterogeneity that exists across respondents. Finally, we compare both deterministic and probabilistic approaches for accommodating the impact of response latency, the latter recognising the fact that the influence of response latency is complicated and in some instances it may have no bearing on neither the utility coefficients nor error variance.

To test our methodology we use a stated choice experiment dataset that was collected via an online survey. This was administered to a pre-recruited panel of individuals residing in Germany and had the aim of establishing their willingness to pay for features present at fishing sites in Denmark. Results from our analysis provide evidence that the variance of the observed factors and preference heterogeneity are sensitive to response latency. As expected, we find that error variance generally decreases with increasing response latency. Importantly, however, we find that under the probabilistic approach that also accommodates the preference heterogeneity the relative decrease is markedly smaller. In fact, results from this model highlight that, irrespective of the length of time respondents take to complete the choice experiment, there is always a proportion associated with relatively high error variance. Accounting for this is found to be highly relevant for estimation outcomes as goodness-of-fit measures are significantly improved. While the estimates of marginal willingness to pay for the fishing site attributes do not appear to be affected, we do observe that the predictions of fishing trips and revenue generated are influenced by response latency.

The remainder of the paper is structured as follows: Section 2 describes our latent class modelling approach; Section 3 outlines our empirical case-study; Section 4 presents the results from the analysis; and, Section 5 concludes.

2 Methods

2.1 Background notation

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:

$$U_{nit} = \beta x_{nit} + \varepsilon_{nit}, \tag{1}$$

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:

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

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

For identification purposes, generally the value of λ is set to unity and, thus, drops out of the probability calculation. However, in cases where it is believed that the variance of the unobserved factors are not the same for all respondents, it is necessary to estimate separate scale parameters for these respondents. In this paper we are interested in exploring whether or not the scale parameter varies with response latencies. Specifically, it is believed 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, in accordance with Haaijer, Kamakura and Wedel (2000), this is likely to be exhibited in the error variance—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. This motivates the present study on how to appropriately identify and accommodate this issue.

2.2 Deterministic approach for accommodating response latency

As a first step to explore the association between response latency and the error variance[2], we specify a deterministic relationship, as follows:

[2]In this paper we model response latency as an independent variable. We remark that response latency could alternatively be considered as a dependent variable. However, this is beyond the scope of this paper. Interested readers are referred to Hess and Stathopoulos (2011) for an approach where response latency is treated as a dependent variable.

$$\lambda_n = \left(\frac{\mathcal{T}_n}{\min(\mathcal{T})}\right)^{\omega}, \tag{3}$$

where $\mathcal{T}$ and ω are respectively the combined time taken to complete all choice tasks and the coefficient relating the response latency with the scale parameter.[3] The rationale for dividing by the minimum response latency is so that the scale parameter takes a value of 1 for the individual(s) who completed in the shortest time, and, thus, the scale parameters for the remaining respondents are estimated relative to 1. Under this representation, a positive value of ω implies that the scale parameter increases (i.e., a reduction in error variance) as response latency lengthens, whereas a negative value of ω would lead to decreasing values of scale (i.e., an increase in error variance) as response latency lengthens.[4]

Although this has appeal, since its directly accommodates any differences that exist in the variance of the unobserved factors due to response latency, it does not address any unobserved preference heterogeneity. In an attempt to also uncover and explain the preference heterogeneity, we, therefore, 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 as well as preferences:

$$\Pr(y_n|x_n) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T_n} \frac{\exp(\lambda_s \beta_s x_{nit})}{\sum_{j=1}^{J} \exp\left(\lambda_s \beta_s x_{njt}\right)}, \tag{4a}$$

where β_s and ω_s are estimated for each of the s classes and where we specify the class probability π_s as follows:

$$\pi_s = \frac{\exp(C_s)}{\sum_{s=1}^{S} \exp(C_s)}, \tag{4b}$$

where C is a class probability constant and where the restriction $C_{s=1} = -\sum_{s=2}^{S} C_s$ is placed so

[3] We 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 subsequent 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 (Malhotra, 2008). Since the time respondents spend on making their choices generally drops as they progress through the experiment (cf. Haaijer, Kamakura and Wedel [2000] and Rose and Black [2006]), this also helps to disentangle the issue from the potential effects of learning and fatigue.

[4] We acknowledge that our deterministic representation of the response latency may not be very flexible. Other deterministic forms (e.g., polynomial representations to accommodate u-shaped and inverse u-shaped forms) were explored, but did not improve model fit. Our decision to move to probabilistic representations also is to accommodate more flexible relationships.

that $\sum_{s=1}^{S} C_s = 0$ for identification purposes. The attraction of such a specification is that, in addition to a finite representation of preferences, we can concurrently assess the influence of response time on the scale parameter within each of the latent segments.

2.3 *Probabilistic approach for accommodating response latency*

Despite the additional insight provided by the specification outlined in Equation (4) over some of the existing approaches used to explore the role of response latency in stated choice experiments, we acknowledge that it is based on a deterministic relationship between response latency and error variance. Consequently, it will only be capable of explaining the general influence of response latency on the variance of the unobserved factors. However, as noted by Bonsall and Lythgoe (2009), response latency is likely to vary between individuals, depending on their personal decision-making styles, and to reflect circumstances, such as the extent of any distraction, the time pressure they are under and their current mental and motivational state. For this reason, there are likely to be instances where the variances of the unobserved factors are not associated with response latency. If this is found to be the case, the deterministic modelling approach could lead to potentially biased and misleading results. Therefore, rather than rely on such a strict specification, a probabilistic approach 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 this paper, we assume that the scale parameter is finitely distributed:

$$\Pr(y_n|x_n) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T_n} \frac{\exp(\mu_s \beta_s x_{nit})}{\sum_{j=1}^{J} \exp\left(\mu_s \beta_s x_{njt}\right)}, \tag{5a}$$

where μ is now used to represent λ to enable them to be distinguished, and where the constraint $\mu_{s=1} = 1$ is placed for identification purposes. Given our focus on response latency and our desire to retrieve probabilistic estimates for the scale parameters at different response latencies we specify π_s as follows:

$$\pi_s = \frac{\exp\left(C_s + \psi_s\left(\frac{\mathcal{T}_n}{\max(\mathcal{T})}\right)\right)}{\sum_{s=1}^{S} \exp\left(C_s + \psi_s\left(\frac{\mathcal{T}_n}{\max(\mathcal{T})}\right)\right)}, \tag{5b}$$

where ψ captures the influence of response latency on class membership, subject to the same constraint $\psi_{s=1} = -\sum_{s=2}^{S} \psi_s$ so that $\sum_{s=1}^{S} \psi_s = 0$. We note that in this case we divide by the maximum response latency, since it conveniently avoids taking exponents of large numbers.

We remark that Equation (5a) also denotes class-specific utility coefficients. This is an

attempt to also facilitate the heterogeneity in preferences that may exist among respondents.[5] This preference heterogeneity can be concurrently accommodated by specifying $S = S_\beta \times S_\mu$ classes, where S_β and S_μ are the number of finite classes explaining preferences and the number of classes within each of these classes with different scale parameters respectively. Constraints need to be placed on the values of β and μ so that they can both be separately identified. Specifically, classes 1 to S_μ are associated with β_1, classes $S_\mu + 1$ to $2S_\mu$ are associated with β_2, classes $2S_\mu + 1$ to $3S_\mu$ are associated with β_3 and so forth, while the values of μ in classes 1, $S_\beta + 1$, $2S_\beta + 1$, and so forth are set to unity for identification.[6]

3 Empirical case-study: German anglers on vacation in Denmark

The conducted stated choice experiment was intended to quantify the value of fishing site characteristics. Specifically, the aim was to uncover German tourists' preferences for recreational fishing at different hypothetical fishing sites when on vacation in Denmark, as German tourists play an important role in the Danish tourism industry. Six different fishing site quality attributes, each with three different levels, were used in the survey. Through extensive focus group testing and a pilot study, these attributes and levels were identified as being of importance for German tourists when angling in Denmark. Other attributes were included in the focus groups and the pilot study, but were either deemed of no importance or unrealistic from the anglers' point of view and were, thus, removed from the final design. Furthermore, a cost attribute with six different levels was included. An overview of all attributes and levels are presented in Table 1.

Table 1 about here.

Using a D-efficiency criterion for evaluation, a Bayesian updated experimental design was employed using priors from a pilot study with 103 respondents. The final experimental design consisted of three blocks, each of which comprised of six choices (which was judged to be the appropriate number of choice tasks following feedback during the focus group discussions). Hence, each respondent evaluated six different choice sets, each of them consisting of two experimentally designed alternatives and a zero-cost 'none-of-these' opt-out alternative.

Respondents were sampled from a pre-recruited internet panel on the condition that they could be classified as recreational anglers since this was considered the relevant target population for the specific survey. Respondents entered a draw for a number of different prizes. A

[5] We note that, while we opt for finite distributions of preferences and scale parameters, the models presented here could be extended to continuous representations. 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.

[6] We recognise the work by Hess and Rose (2012) highlighting the difficulty in separating preference and scale heterogeneity. While the probabilistic approach does afford a more flexible model that can go some way towards separating the two, we remark that care is needed when interpreting the results.

total of 1,006 respondents replied to the questionnaire, out of which 156 questionnaires were uncompleted, thus leaving an effective sample of 850 respondents and 5,100 choice observations for the following analysis. The socio-demographic distribution of these respondents is as follows: 56 percent are male, the average age is 37 years, 30 percent have an university degree, and the average annual household income is in the region of €41,000.

Response latency for each respondent was measured by automatically logging the time taken between clicking the 'next' button at the bottom of the survey webpage. This recording of response latency was unobservable to the respondent, and as such no extra effort was required by the respondent for us to retrieve this paradata. For the choice experiment section of the questionnaire, each choice task was presented on a separate webpage and respondents were first asked to mark their chosen alternative in the choice set and then click the 'next' button to proceed to the next choice set in the sequence (thus, enabling respondents the opportunity to amend their decision before progressing). While there is appeal in investigating the impact of response latency at the choice task level, for the reasons mentioned in Footnote 3, we use response latency for the entire choice task sequence including the scenario description.

4 Results and discussion

4.1 Estimation results

4.1.1 Initial estimation results

Results from the basic MNL model (Equation (2)) presented in Table 2 show that all but one attribute coefficient are significant. The insignificant parameter suggests that German anglers are indifferent between a stream with a mix of smaller and larger fish and a stream with mainly smaller fish (and a few large ones). The signs of the estimated coefficients are as expected—with positive signs for 'Catch opportunity', 'Fish size', 'Nature experience', and 'Quality of the fishing water' and negative signs associated with 'Distance from accommodation to site', 'Number of other anglers at site', and the 'Cost' attribute. Moreover, the results comply with our prior expectations with regard to being monotonic, except for the attribute 'Number of other anglers at site'. Although not significant, the latter suggests that German anglers do not obtain any additional dis-utility, when there are 'many' other anglers present at the site compared to a situation where there are only 'few'. Finally, the results show, that the ASC for the status-quo alternative is negative and significant, implying that the anglers, all else equal, prefer one of the two generic alternatives instead of opting for the 'none of these' alternative.

Table 2 about here.

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)[7] 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 −9.991 (which represents an average predicted probability of 0.19), whereas the highest is −4.011 (representing an average predicted probability of 0.51). The mean and median are found to be −6.056 and −5.927 respectively.

Figure 1 about here.

Moving our attention to the histogram associated with latency, we find that the median is 77 seconds and that most respondents (just over two-thirds) required between 1–2 minutes to complete the entire choice experiment (which equates to an average of 10–20 seconds per choice task). While 90 percent of respondents completed the choice experiment within 3 minutes, a few spent almost 1 hour completing the exercise (i.e., almost an average of 10 minutes per choice task). While this could of course be interpreted as respondents spending a huge effort on answering the six choice sets, we find it more reasonable to suspect that this is rather due to measurement error in terms of respondents not focusing only on the choice experiment in that period of time. 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 answering the choice experiment. At the other end of the spectrum a number of respondents completed the six choice tasks in less than 15 seconds (which is equivalent to an average of 2.5 seconds per task). This is quite surprising considering that this includes the time required to load the webpage, which was likely to be in the region of 1–2 seconds[8]. While it is clearly not trivial to determine exactly what the minimum required response latency would be in order to obtain preference elicitations, response times of 2.5 seconds per choice task would, *prima facie*, seem to be very short. It is also of concern that these extremely 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.

[7]Each respondent's contribution to the MNL LL can be retrieved using $\mathrm{LL}_n = \sum_{t_n=1}^{T_n} \ln\left(\Pr\left(i_{nt}|\hat{\beta}, x_n\right)\right)$.

[8]Ideally, the time it takes for the webpage to load should be subtracted from the response latency measure. Unfortunately, exact load times for the specific survey webpage were not measured. Load time would, of course, depend on the survey webpage itself but also on the respondent's internet connection. Based on the current average internet connection speed in Germany of 8 mbps (SpeedTests.net, 2011), we conjecture that our respondents typically experienced load times in the region of 1-2 seconds.

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 number of cases where the respondent's contribution to log-likelihood falls below the null log-likelihood associated with a random sequence of choices (i.e., $\ln(1/3) \times 6$). These findings motivate our search for more flexible models and to explore the potential of using the response latencies to help describe the heterogeneity in error variances and preferences.

4.1.2 Deterministic estimation results

We begin this analysis under the assumption that response latency has a deterministic influence on the unobserved error variance. Results obtained from two such models are presented in Table 3. In the first of these models (labelled Det1) the scale parameter is reparameterised in accordance with Equation (3). We remark that we find similar findings regarding sign and significance as well as monotonicity as those obtained under the MNL model. Importantly, the coefficient linked with the scale parameter is positive as well as significant. The fact that it is positive suggests that scale increases, and, hence, error variance decreases, as response latency increases. This conforms with our a priori expectations and evidence presented elsewhere (Haaijer, Kamakura and Wedel, 2000). Since the coefficient is less than 1, we can also say that the influence of latency on the scale parameter diminishes as response time increases, which we note is a further anticipated finding. As one moves from the MNL model to Det1 we observe an improvement of 56 log-likelihood units at the expense of fitting one additional parameter. The null embedded in the MNL model provides a likelihood ratio test statistic of 112.75 against the χ^2 critical value of 3.84 $\left(\chi^2_{1,0.05}\right)$.

Table 3 about here.

Notwithstanding the additional insight provided by Det1, it is based on the assumption that all respondents have the same preferences. As this is usually a very strict and in many cases an unrealistic assumption we introduce a second model (labelled Det2). This is based on Equation (4) and assumes that the heterogeneity in preferences can be adequately captured using two classes.[9] An observation of the estimates produced from this model reveals that the attribute coefficients under both classes have the same sign as the previous models, although

[9] We note that our analysis in this paper did not follow a conventional latent class model specification test to establish the appropriate number of classes. Our intention in this paper is primarily illustrative, but we acknowledge that there is scope to facilitate more classes and encourage others to do so.

we do remark some differences relating to significance of these parameters. In terms of implied ranking of attributes and significance of attribute coefficients, class 1 has similarities to those observed in previous models. In contrast, anglers belonging in class 2 appear to only have significant preferences for characteristics of the site and its surroundings, namely 'Quality of the fishing water', 'Distance from accommodation to site' and the highest level of 'Nature experience', rather than the more fish specific characteristics, namely 'Catch opportunity' and 'Fish size'. A further important distinction is the positive ASC estimated for class 2, which would indicate, all else being equal, that these anglers favour opting-out of the fishing experience whereas anglers in class 1 would rather not opt-out. In summary, the model thus suggests that there is one class of anglers who would prefer to go fishing regardless of the type of fishing experience described, but their utility is affected by all aspects of the fishing site. The other class of anglers would rather not go fishing unless the natural and environmental qualities of the fishing site are high but appear indifferent about the fish attributes in terms of numbers and sizes of potential catch.

Of great interest are the estimated ω terms. Results from the first class show that the value of ω is positive and significant, which conforms to findings in Det1. However, the ω term for the second class is negative as well as significant. While, on the face of it, this would appear as a somewhat unexpected result, since it implies a larger and increasing error variance with increasing response latency, the value of the class probability constant in class 2 highlights that this is the minority class of approximately 32 percent. This is an important finding. This highlights that, while for the majority of anglers an increased response latency is associated with a higher scale parameter, for a smaller subset it is connected with a lowering of the scale parameter. 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. Putting this finding together with the class-specific utility coefficients provides us with a richer insight, which supports our move to this more flexible specification. Finally, compared to Det1, the model fit achieved under Det2 is much superior. While, we acknowledge that this improvement reflects the fact that it takes the panel nature of the data into account, we note that there is a reduction of over 600 log-likelihood units. This comes at the expense of 16 additional parameters, which contributes to a significant likelihood ratio test statistic.

Notwithstanding the improvements in model fit witnessed under these models that assume a deterministic relationship, the influence of response latency may be ambiguous. Indeed, there are likely to be instances where the variances and preferences associated with quick responses are not dissimilar to those relating to longer latencies. The use of a probabilistic approach may, therefore, be justified.

4.1.3 Probabilistic estimation results

Table 4 presents the results from two separate models, derived from Equation (5), which attempt to probabilistically identify the impact of response latency. The first specification (labelled Prb1) is based on the assumptions of preference homogeneity among all respondents and that the distribution in the scale parameter can be sufficiently described using two finite points. Focusing firstly on the utility coefficients obtained for this model, we note that they are in line with those uncovered under the MNL model. Interestingly, however, we note that all attribute coefficients are significant and the monotonicity in the intensity of the attribute levels is respected in all cases. Moving to the scale parameter estimated for the second class, reveals that it is estimated with a value less than 0.001. This, along with the fact that it is not significantly different from zero, effectively implies that a subset of respondents choose completely randomly. We further note that it is significantly different from one, indicating that the two classes can be distinguished according to error variance. The estimated values of the class membership constants indicate that, other things being equal (i.e., at the shortest response latency), approximately 62 percent of respondents are predicted as belonging in the class associated with making random choices. While this confirms our suspicion that the very fast choices tend to introduce more noise in our models, the other side of the coin implies that a predicted 38 percent of respondents at the shortest response latency are making non-random choices. The fact that the value of ψ for the second class is negative implies that the proportion identified as having made completely random choices decreases as response latency increases—which is consistent with our earlier results. In fact, calculating this at the median response latency of 77 seconds reveals that this probability drops to almost 40 percent and is practically zero for those with a response latency of over 8 minutes.[10] Comparing Prb1 to the Det1 model, we observe an increase in model fit by 86 units, at the cost of just two additional parameters.

Table 4 about here.

So far our models have accounted for the link between response latency and the variance of the unobserved factors. But, as suggested by Rose and Black (2006), it is also conceivable that response time may be related with taste intensities. To establish if this is the case, our second probabilistic model (labelled Prb2) assumes that the distribution of preferences can be accommodated using two latent segments (i.e., $S_\beta = 2$) and that each segment can be further partitioned into two groups with different error variance (i.e., $S_\mu = 2$). This leads to a 4-class model, whereby classes 1 and 2 are associated with β_1 and classes 3 and 4 are associated with β_2 as well as the constraint that the values of μ in classes 1 and 3 are equal to 1. Looking firstly at the utility coefficients obtained for classes 1 and 2, reveals that they

[10] See Appendix A.

share many similarities with those retrieved under the Prb1 model, the only exception being that two of the coefficients are not significant. The utility coefficients associated with classes 3 and 4 highlight that, aside from their magnitude (which is an artefact of the difference in scales), they are comparable to those uncovered in the second class in the Det2 model, but with more of them estimated as being significant. Under the Prb2 model, the scale parameters associated with classes 2 and 4 are both found to be significantly different from zero, but more importantly also different from 1. As a result, we find that the further segmentation of the preference classes is warranted and in both cases (i.e., classes 2 and 4) they have a higher error variance. We remark that the relative difference in error variance is much larger for the subgroup associated with the second set of taste intensities. The magnitude and sign of the class membership constants identifies that respondents are most likely to be in class 2, followed by class 3, at the shortest response latency. In fact, the estimates suggest that over 95 percent of respondents with this response latency belong in these classes (76 percent and 20 percent respectively in classes 2 and 3). Interestingly, response latency is found to be a significant factor in explaining the class probabilities. While difficult to interpret, these suggest that, all else held constant, as response latency increases the membership of classes 1 and 4 increase. We remark that at the median response latency, over half of respondents are predicted to belong in class 1, with most of the remaining respondents predicted in class 4. Moreover, class membership probabilities stabilise at response latencies of 2 minutes and above at around 60 percent and 40 percent in classes 1 and 4 respectively, indicating that irrespective of the length of time respondents take to complete the choice tasks there is always a proportion associated with relatively high error variances (class 4). We can also see that the more flexible specification, which probabilistically incorporates the role that response latency has on error variance and taste intensities, leads to a better model fit. We note that this improvement in fit is supported by the $\bar{\rho}^2$, AIC and BIC statistics, even after accounting for the additional parameters.

4.1.4 Response latency and relative variance

Findings presented in Tables 3 and 4 indicate that response latency has a bearing on the variance of the unobserved factors. To illustrate this, in Figure 2 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. We further note that the relative variances connected with the Prb1 and Prb2 models are calculated on the basis of the unconditional class membership probabilities (and, thus, represent the most likely relative variance at the given latency).

Figure 2 about here.

Inspecting Figure 2 reveals that, with the exception of the MNL model (which assumes homoscedasticity, irrespective of response latency), there is 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 Det1, which suggests that the variance associated with respondents who required 1 minute to complete the choice experiment is approximately one-fifth of that associated with those who completed the experiment in the shortest time. However, we note that this model accounted for neither preference heterogeneity nor a probabilistic classification of the variance, 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 Det2 and Prb1 models. Under these models, the variance of the unobserved factors for respondents who completed the experiment in the region of 2 minutes is (on average, in the case of the Det2 model) approximately half 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 Prb2, we see a somewhat different pattern. While the error variance is shown to decrease in accordance with the previous models, the reduction is to a lesser extent. In fact, under this model, the error variance is, on average, approximately only 10 percent lower for respondents who required at least 1 minute relative to those who answered in the quickest time. The fact that this differs so much from the previous models highlights the importance of accommodating scale and preferences in a probabilistic manner, otherwise any inferences regarding response latency and error variance may be misleading.

4.2 Willingness to pay and demand analysis

The comparison of parameter estimates between the various models is not possible since they are each subject to a different scaling of the parameter estimates. What does make comparative sense, and which are of potentially greater interest to policy analysts, are the marginal willingness to pay estimates, since the scale effect is neutralised. In Table 5 we report these estimates for all attributes and across all model specifications. To enable straightforward comparison of significance, we also report their associated confidence intervals (estimated using the delta method). Although model fits have been shown to be different under each of the specifications, the confidence intervals of the willingness to pay estimates retrieved under each model specification all overlap. Further inspection using t-tests also reveals the differences by and large are not significant. Nevertheless, the models which accounted for preference heterogeneity (Det2 and Prb2) have, for the most part, produced slightly, albeit not significantly, lower (in absolute terms) estimates.

Table 5 about here.

The relatively minor changes in the willingness to pay estimates and the absence of significant differences is not an entirely surprising finding, given that the differences in error variance are neutralised in the calculation. Nevertheless, due to the differences in scale parameters there is a consequence for probability predictions at different response latencies, which makes it relevant for forecasting fishing trips. To further tease out the effects of response latency, we, therefore, use the model estimates to simulate fishing trips made by German anglers. For this analysis, we consider that their choice of fishing sites is restricted to two options:

Site 1: high catch opportunity, large fish and the possibility of record-breaking large fish, big nature experience, clean and clear fishing water quality, located 4–20 km from accommodation, few other anglers and one-day fishing license costing €40.

Site 2: medium catch opportunity, a mix of smaller and larger fish, varied nature experience, reasonably clean and clear fishing water quality, located more than 20 km from accommodation, many other anglers and one-day fishing license costing €75.

We remark that site 1, according to attribute parameter estimates obtained in the previous models, can be considered a dominant choice. Thus, it would be expected to be associated with a much higher probability prediction compared to Site 2. However, in cases where the variance is high (e.g., short response latencies), the probabilities will be quite comparable.

Using the coefficients uncovered in estimation the probabilities associated with these two sites can be predicted along with the probability of opting out. In Figures 3(a) and 3(b) we plot the means of probabilities (i.e., market shares, or fishing trips) relating to sites 1 and 2 respectively. As expected, the first site is predicted to receive more anglers than the second site. But the increased predicted share is shown to depend on model specification. Importantly, with the exception of the MNL model, where predictions of fishing trips are constant, we also observe that the predictions are sensitive to response latency. Surprisingly, given that site 2 is an inferior choice, share predictions are initially relatively high at 15-25 percent for this site. However, as response latency increases these predictions fall to below 10 percent for all models. Thus, we can infer that the predictions are an artefact of the relatively higher error variance linked with answering quickly. Furthermore, it is interesting to note that the models accounting for preference heterogeneity (Det2 and Prb2) predict fewer visits to site 2 at short response latencies than their preference homogeneity assuming counterparts. Furthermore, they also reach their minimum predicted shares at smaller response times, both of the models stabilising just around the median latency of 77 seconds. In comparison, the Det1 and Prb1 models predict higher shares at the median response time and do not fall below the 10 percent share prediction until a response latency of almost 3 minutes, which is associated with the ninth decile. Finally, looking at the predictions of trips to site 1 and site 2 in Figures 3(a) and 3(b) it is interesting to note that the models that do not take heterogeneity into account tend to under predict the share of opt-out choices.

Figure 3 about here.

Taking the trip predictions and the daily license fee at each site into account it is possible to derive the average expected total revenue per angler per fishing day by offering German anglers on holiday in Denmark a choice between these two fishing sites. Comparing these in Figure 3(c) reveals some discrepancies between the models. While the MNL model, of course, predicts a constant revenue regardless of response latency, the predicted revenue increases with increasing response latency for three of the more flexible models. This is seen especially for the Det1 model, where the model at the shortest response latency would predict an expected revenue of less than €30 per angler per fishing day while at the highest response latency the predicted expected revenue would be almost €40. For the remaining models the impact of response latency is less drastic, and for the Prb2 model the average predicted expected revenue is even decreasing until a response latency of approximately 1 minute, after which it increases and stabilises at around €32. We also draw attention to the potential repercussions of failing to account for preference heterogeneity and, in particular, not accounting for the differences in the opt-out ASC. Focusing at the median response latency of 77 seconds for the current sample in Figure 3(c) it is evident that the models accommodating heterogeneous preferences predict an expected revenue that is €4–7 lower per angler per fishing day than the models assuming homogeneity of preferences. Considering the fact that the two models accommodating for heterogeneity are superior to their counterparts in terms of model fit and, thus, provide a far better description of the data, we conclude that expected revenue is likely to be overestimated if preference heterogeneity is not taken into account. This is mainly due to the fact that the share of anglers opting out is underestimated. While different simulations would be observed under different scenarios, the results, nevertheless, do provide an indication that predictions vary with response latency. Moreover, given that the number of fishing trips made by German anglers per year in Denmark is estimated at 1.1 million[11], the differences in expected fishing license revenue could be in the range of €4.4–7.7 million per year for the given example.

5 Conclusion

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

[11] Estimate based on the predicted 3.1 million overnight stays made by German tourists who engaged in recreational angling when on vacation in Denmark in 2008 (Jensen et al., 2010). This number is multiplied with an estimate for the average number of fishing trips per overnight stay as calculated based on supplementary questions in the questionnaire concerning respondents' number of fishing trips and number of overnight stays during their last vacation in Denmark.

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 and error variance. Specifically, we investigate the issue using a number of scale-adjusted latent class models, where the influence of response latency is tested under both deterministic and probabilistic assumptions. We test our methodology using an empirical dataset collected via an online survey to establish the value that German anglers are willing to pay for fishing site attributes in Denmark.

Results from our analysis reinforce previous findings (e.g., Haaijer, Kamakura and Wedel (2000)) that there is a negative correlation between response latency and error variance. While we observe this under all our specifications, the rate and degree of this reduction is dependent on the model specification. Our probabilistic models seem better equipped to deal with the ambiguous relationship between response latency and error variance. Indeed, we find that the error variance of the observed factors was significantly larger for a subset of the quickest respondents. While this proportion is shown to fall as response latency increases, we also learn that there is always a share of respondents with relatively large error variance. Moreover, in the context of exploring the impact of response latency on error variance, we highlight the importance of also accounting for preference heterogeneity—accounting for only its influence on error variance is found to lead to biased results. Importantly, in accordance with earlier studies (e.g., Haaijer, Kamakura and Wedel [2000], Rose and Black [2006], Otter, Allenby and van Zandt [2008] and Vista, Rosenberger and Collins [2009]) we show that accounting for the role of response latency is important for goodness-of-fit measures. In a similar vein as the inferences made by Hess and Stathopoulos (2011), we further note that specifications assuming a probabilistic relationship and those that facilitate preference heterogeneity outperform their counterparts that are based on the assumption of a deterministic relationship and those derived from homogeneous preferences respectively.

Methodological findings aside, we find that German anglers are estimated with highest willingness to pay values for the attribute relating to ‘Quality of the fishing site’, followed by ‘Catch opportunity’, ‘Nature experience’ and ‘Fish size’, whereas they place negative values on the ‘Distance from accommodation to site’ and ‘Number of other anglers at site’ attributes. While these willingness to pay estimates are found to be relatively comparable

across the various model specifications, we find that response latency should be an important consideration when forecasting sites that the anglers will visit. We, subsequently, show that this has implications when estimating expected revenue. In particular, we show that revenues could be seriously overestimated unless response latency is accounted for.

From data quality and, as we have seen, from forecasting 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 discrete choice 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 Algers [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 an environmental 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 and error variance, not to mention the potential presence of attribute non-attendance or other simplifying heuristic decision rules. 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).

Finally, 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., MacMillan, Hanley and Lienhoop [2006] and Cook et al. [2011]), 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.

References

Bonsall, P. and Lythgoe, B. (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 Algers, S. (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., Kingsley, D., Peterson, G. L., Flores, N. E., Clarke, A. and Birjulin, A. (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., Hensher, D. A. and Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: a latent class specification, *Journal of Environmental Planning and Management* **54**(8): 1061–1076.

Cook, J., Jeuland, M., Maskery, B. and Whittington, D. (2011). Giving stated preference respondents "time to think": results from four countries, *Environmental and Resource Economics* pp. 1–24.

Fleming, C. M. and Bowden, M. (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, *MAGKS Papers on Economics 201002*, Philipps-Universitt Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung). **URL:** *http://ideas.repec.org/p/mar/magkse/201002.html*

Haaijer, R., Kamakura, W. and Wedel, M. (2000). Response latencies in the analysis of conjoint choice experiments, *Journal of Marketing Research* **37**(3): 376–382.

Hess, S. and Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models?, *Transportation* **39**(6): 1225–1239.

Hess, S. and Stathopoulos, A. (2011). Linking response quality to survey engagement: a combined random scale and latent variable approach, *Journal of Choice Modelling*, forthcoming.

Holmes, T., Alger, K., Zinkhan, C. and Mercer, E. (1998). The effect of response time on conjoint analysis estimates of rainforest protection values, *Journal of Forest Economics* **4**(1): 7–28.

Jensen, C. L., Nissen, C. J., Olsen, S. B. and Boesen, M. (2010). An analysis of German anglers' choice of country for vacation - focusing on Denmark (in Danish), *Working paper 1/2010*, Institute of Food and Resource Economics, University of Copenhagen.

Lindhjem, H. and Navrud, S. (2011a). Are internet surveys an alternative to face-to-face interviews in contingent valuation?, *Ecological Economics* **70**(9): 1628–1637.

Lindhjem, H. and Navrud, S. (2011b). Using internet in stated preference surveys: a review and comparison of survey modes, *International Review of Environmental and Resource Economics* **5**: 309–351.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*, New York: Oxford University Press.

MacMillan, D., Hanley, N. and Lienhoop, N. (2006). Contingent valuation: environmental polling or preference engine?, *Ecological Economics* **60**(1): 299–307.

Magidson, J. and Vermunt, J. K. (2008). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference, *Sawtooth Software Conference*.

Malhotra, N. (2008). Completion time and response order effects in web surveys, *Public opinion quarterly* **72**(5): 914.

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., Allenby, G. M. and van Zandt, T. (2008). An integrated model of discrete choice and response time, *Journal of Marketing Research* **45**(5): 593–607.

Rose, J. and Black, I. (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.

Schwappach, D. L. B. and Strasmann, T. J. (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.

SpeedTests.net (2011). Germany internet speed statistics: download and upload test results. **URL:** *http://www.speedtests.net/world/de/, October.*

Vista, A. B., Rosenberger, R. S. and Collins, A. R. (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.

Appendix A: Class membership probabilities for the probabilistic models

Using Equation (5b) and the coefficients estimated under the Prb1 and Prb2 models in Table 4 the (unconditional) class membership probabilities can be derived for any given response latency. In Figures A1(a) and A1(b) respectively we plot the results of these calculations across the range of response latencies.

Figure A1 about here.

For Prb1 the graphical illustration shows that for the shortest response latency there is a 62 percent probability of belonging to the class where scale is almost zero, which would be associated with completely random choices. As response latency is increased, the class membership probability for this class reduces quite linearly with 10 percentage points per minute (notice that the horizontal axis is scaled logarithmically for illustrative purposes) until at response latencies above 10 minutes effectively reaching zero. This being a two-class model, the membership probability for the class where scale is normalised to unity is of course negatively proportionate to this. Not surprisingly, this highlights that as response latency is increased, the proportion of respondents predicted as not choosing randomly increases. It is, however, of some concern that this model would also suggest that at the median response latency for this sample almost 50 percent of the choices are predicted as having effectively been made randomly.

For the Prb2 model, the largest class (with a class membership probability of 0.76) at the shortest response latency is class 2 where the model picks up a specific preference structure though with a lot of noise relative to others with the same preference structure in class 1 (where the probability is approximately 2 percent). At this short response latency there is furthermore a 20 percent probability of belonging to class 3 which holds a somewhat different preference structure than classes 1 and 2, while only a small proportion of around 2 percent would be expected to end up in class 4 where choices would appear to have been very random. As response latency increases beyond 2 minutes, the probability of obtaining relatively high error variance choices stabilises at around 40 percent. While still high, it is somewhat more reassuring than the 50 percent suggested by Prb1, which further highlights the necessity to also accommodate for the heterogeneity in preferences and the more flexible means of dealing with the differences in scale.

Table 1: Attributes and levels used in the choice experiment

Attributes	Levels (coding)
Catch opportunity	Low (Catch0) Medium (Catch1) High (Catch2)
Fish size	Many small and few large fish (Size0) A mix of smaller and larger fish (Size1) Large fish and the possibility of record-breaking large fish (Size2)
Nature experience	Small (Nature0) Varied (Nature1) Big (Nature2)
Quality of the fishing water	Unclear and appears to be polluted (Quality0) Reasonably clean and clear (Quality1) Clean and clear (Quality2)
Distance from accommodation to site	<4 km (Distance0) 4–20 km (Distance1) >20 km (Distance2)
Number of other anglers at site	None (Number0) Few (Number1) Many (Number2)
Cost of one-day fishing license	€7, €15, €25, €40, €75 and €140 (Cost)

Table 2: Estimation results (MNL model)

	est.	$\lvert t\text{-rat.}\rvert$
β_{Catch1}	0.281	5.60
β_{Catch2}	0.492	9.69
β_{Size1}	0.025	0.50
β_{Size2}	0.142	2.68
β_{Nature1}	0.214	4.01
β_{Nature2}	0.483	10.23
β_{Quality1}	0.392	6.87
β_{Quality2}	0.686	12.84
$\beta_{\text{Distance1}}$	-0.196	3.93
$\beta_{\text{Distance2}}$	-0.351	6.33
β_{Number1}	-0.306	5.46
β_{Number2}	-0.250	4.76
β_{Cost}	-0.011	20.28
$\beta_{\text{ASC(SQ)}}$	-0.242	2.84
$\mathcal{L}\left(\hat{\beta}\right)$	-5,147.779	
K	14	
$\bar{\rho}^2$	0.079	
AIC/N	2.024	
BIC/N	2.023	

Table 3: Estimation results (deterministic scale adjusted models)

	Det1		Det2			
			Class 1		Class 2	
	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
β_{Catch1}	0.123	5.06	0.212	5.53	0.810	0.59
β_{Catch2}	0.227	7.91	0.343	8.01	1.878	1.49
β_{Size1}	0.020	0.92	0.052	1.50	0.048	0.04
β_{Size2}	0.081	3.42	0.113	2.89	2.715	1.77
β_{Nature1}	0.097	3.91	0.099	2.59	0.000	0.00
β_{Nature2}	0.225	8.16	0.298	7.63	4.165	3.06
β_{Quality1}	0.201	6.87	0.294	6.69	4.108	2.49
β_{Quality2}	0.331	9.46	0.464	9.04	4.617	3.30
$\beta_{\text{Distance1}}$	-0.084	3.66	-0.079	2.19	-2.365	1.98
$\beta_{\text{Distance2}}$	-0.158	5.82	-0.215	4.83	-2.266	1.97
β_{Number1}	-0.131	4.86	-0.177	3.91	-0.473	0.43
β_{Number2}	-0.126	5.07	-0.189	4.63	-1.276	1.10
β_{Cost}	-0.005	10.61	-0.007	10.15	-0.346	3.70
$\beta_{\text{ASC(SQ)}}$	-0.085	2.26	-0.771	7.86	4.356	2.27
ω	0.372	10.98	0.270	7.07	-0.887	8.22
C	-	-	0.380	8.11	-0.380	8.11
$\mathcal{L}(\hat{\beta})$	-5,091.404		-4,485.503			
K	15		31			
$\bar{\rho}^2$	0.089		0.194			
AIC/N	2.003		1.771			
BIC/N	2.001		1.767			

Table 4: Estimation results (probabilistic scale adjusted models)

	Prb1		Prb2			
	Classes 1–2		Classes 1–2		Classes 3–4	
	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
β_{Catch1}	0.449	5.51	0.458	5.63	205.142	1.41
β_{Catch2}	0.987	11.37	0.764	8.87	227.057	4.45
β_{Size1}	0.309	3.80	0.117	1.47	117.554	0.97
β_{Size2}	0.493	5.58	0.256	2.87	554.075	7.06
β_{Nature1}	0.191	2.17	0.228	2.64	65.327	2.04
β_{Nature2}	0.696	9.74	0.673	8.84	770.238	10.14
β_{Quality1}	0.914	8.98	0.701	6.92	943.975	5.80
β_{Quality2}	1.389	13.51	1.035	10.12	892.030	9.82
$\beta_{\text{Distance1}}$	-0.261	3.44	-0.123	1.56	-405.086	6.89
$\beta_{\text{Distance2}}$	-0.668	6.81	-0.468	4.79	-137.414	0.85
β_{Number1}	-0.348	3.90	-0.375	3.87	-377.511	2.80
β_{Number2}	-0.420	4.74	-0.428	4.82	-736.641	25.91
β_{Cost}	-0.021	16.61	-0.015	12.10	-42.090	5.95
$\beta_{\text{ASC(SQ)}}$	-0.943	5.57	-2.025	9.04	709.670	19.31
μ	0.000	0.02	0.449	9.07	0.001	7.10
$C_{\mu=1}$	-0.241	2.04	-1.577	4.93	1.045	3.12
$C_{\mu\neq1}$	0.241	2.04	2.436	6.01	-1.904	5.93
$\psi_{\mu=1}$	13.908	3.08	105.099	5.61	-92.428	4.49
$\psi_{\mu\neq1}$	-13.908	3.08	-117.695	3.86	105.025	5.60
$\mathcal{L}\left(\hat{\beta}\right)$	-5,005.324		-4,415.462			
K	17		36			
$\bar{\rho}^2$	0.104		0.206			
AIC/N	1.970		1.746			
BIC/N	1.968		1.740			

Table 5: Marginal willingness to pay estimates (€ per day (confidence intervals in parentheses))

	MNL	Det1	Det2[a]	Prb1	Prb2[ab]
Catch1	24.92 (16.39,33.46)	23.59 (15.55,31.63)	22.29 (15.36,29.22)	21.82 (14.13,29.51)	20.04 (10.06,30.02)
Catch2	43.70 (34.52,52.87)	43.56 (35.00,52.12)	36.62 (29.28,43.97)	47.98 (40.60,55.36)	32.28 (21.89,42.66)
Size1	2.21 (-6.52,10.93)	3.84 (-4.34,12.02)	5.34 (-1.71,12.40)	15.04 (7.59,22.48)	5.76 (-4.50,16.02)
Size2	12.64 (3.47,21.81)	15.50 (6.85,24.14)	14.03 (6.26,21.80)	23.95 (15.82,32.08)	15.55 (4.19,27.00)
Nature1	19.00 (9.37,28.64)	18.60 (9.53,27.67)	10.11 (2.38,17.84)	9.30 (0.55,18.04)	9.60 (-2.39,21.59)
Nature2	42.90 (33.73,52.07)	43.11 (34.51,51.70)	34.12 (26.77,41.48)	33.84 (26.19,41.50)	34.12 (22.95,45.30)
Quality1	34.80 (24.49,45.11)	38.48 (28.68,48.28)	33.72 (25.41,42.02)	44.42 (35.19,53.64)	36.94 (24.68,49.21)
Quality2	60.90 (50.69,71.12)	63.43 (53.75,73.11)	51.37 (42.95,59.78)	67.49 (58.39,76.58)	49.53 (37.48,61.58)
Distance1	-17.43 (-26.26,-8.60)	-16.00 (-24.26,-7.74)	-10.20 (-17.38,-3.02)	-12.68 (-20.08,-5.28)	-8.85 (-19.35,1.65)
Distance2	-31.16 (-40.50,-21.81)	-30.23 (-39.00,-21.46)	-23.92 (-31.74,-16.10)	-32.44 (-41.08,-23.79)	-19.77 (-31.46,-8.09)
Number1	-27.14 (-37.46,-16.81)	-25.15 (-34.84,-15.46)	-18.41 (-27.46,-9.36)	-16.92 (-25.96,-7.87)	-18.48 (-32.49,-4.46)
Number2	-22.21 (-31.53,-12.90)	-24.06 (-32.86,-15.25)	-20.40 (-28.39,-12.42)	-20.42 (-28.95,-11.90)	-24.12 (-36.17,-12.08)

[a]Reported estimates are weighted on the basis of the class membership probabilities.

[b]Calculations based on median response latency of 77 seconds.

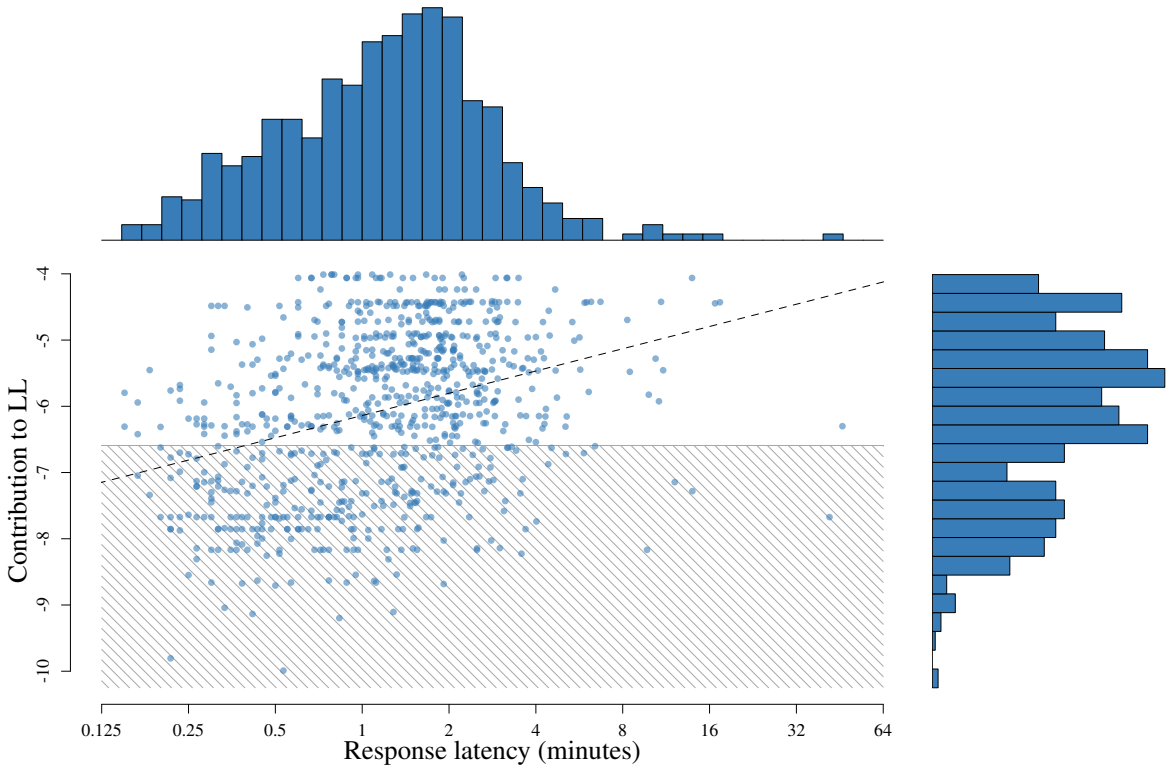


Figure 1: Response latency versus MNL contribution to log-likelihood

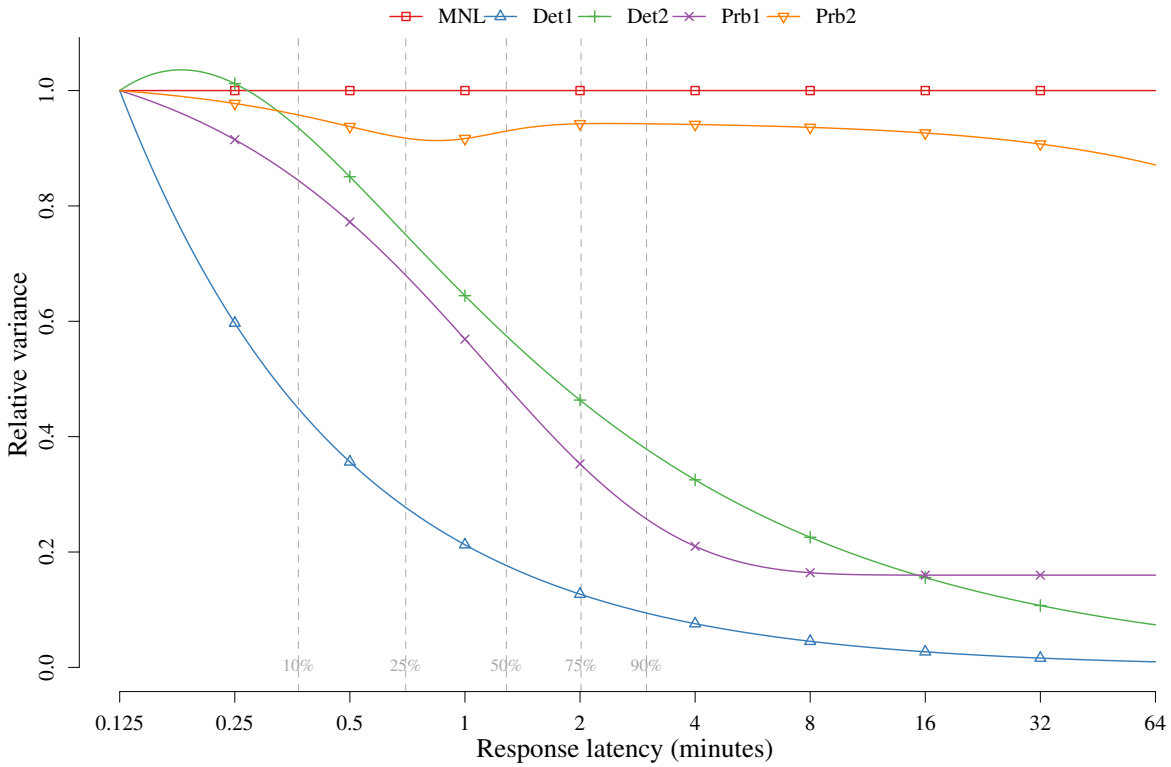
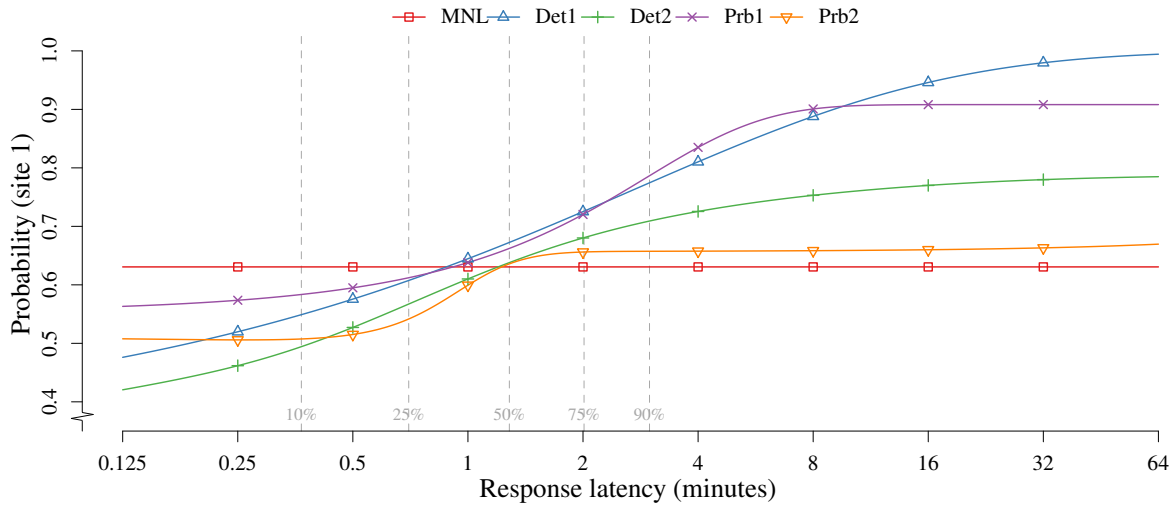
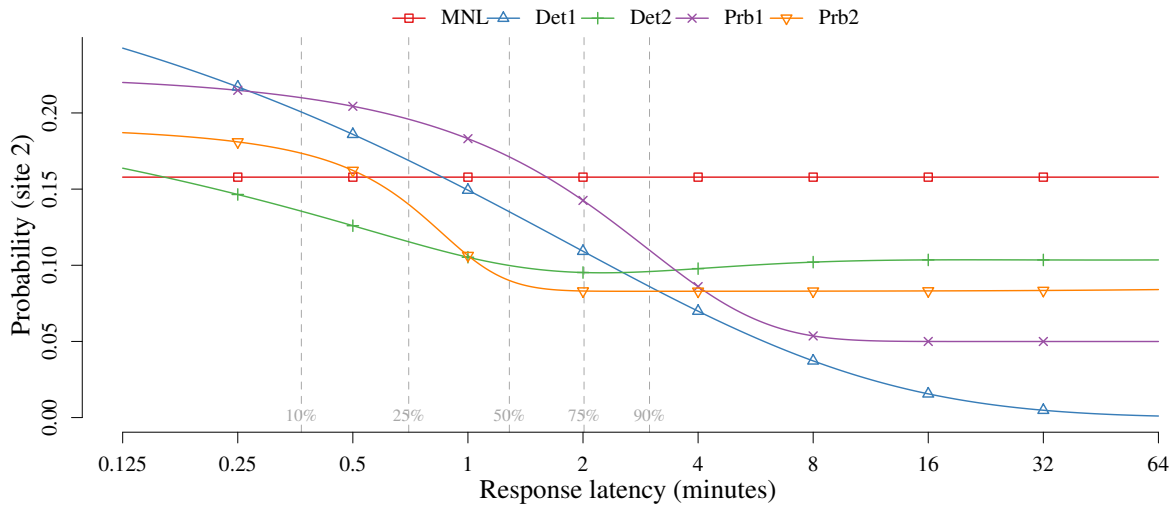


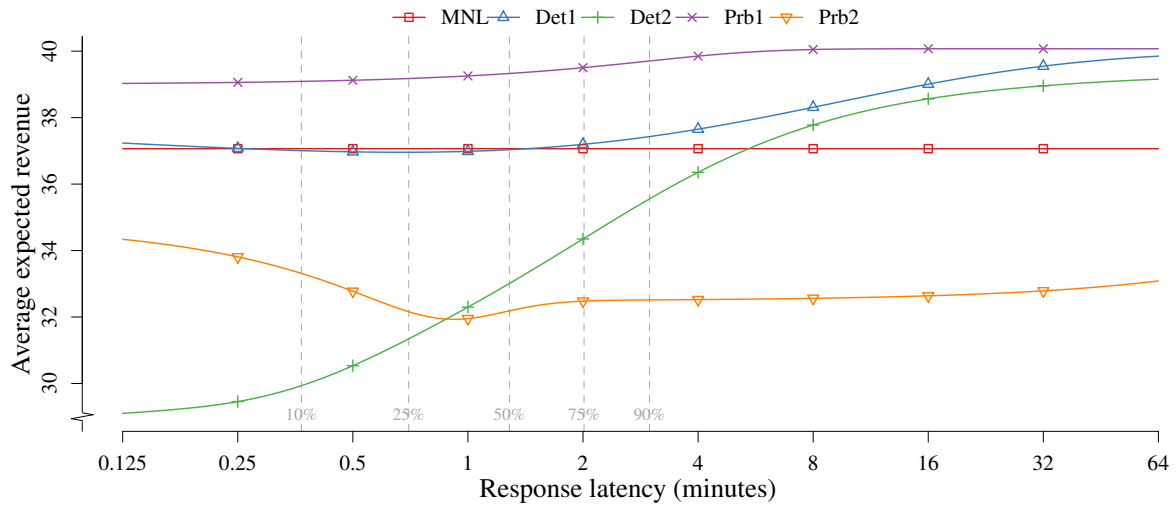
Figure 2: Response latency versus relative variance across all models



(a) Response latency versus average probability of site 1

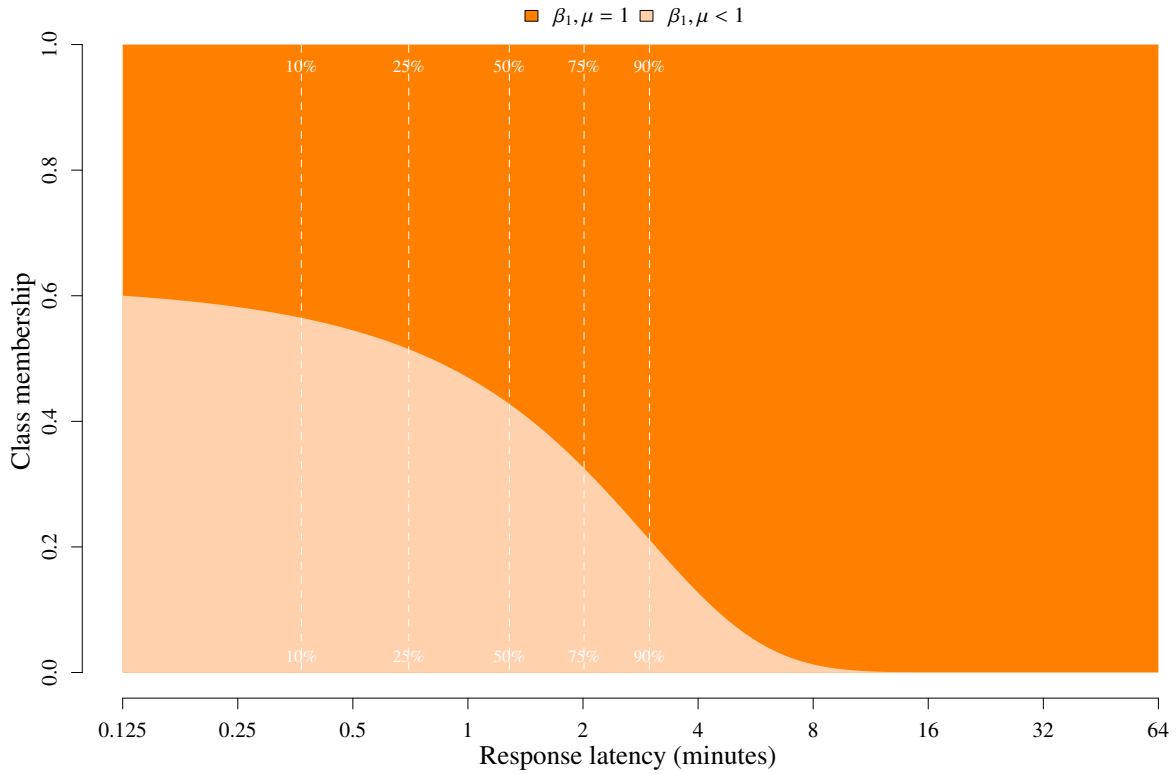


(b) Response latency versus average probability of site 2

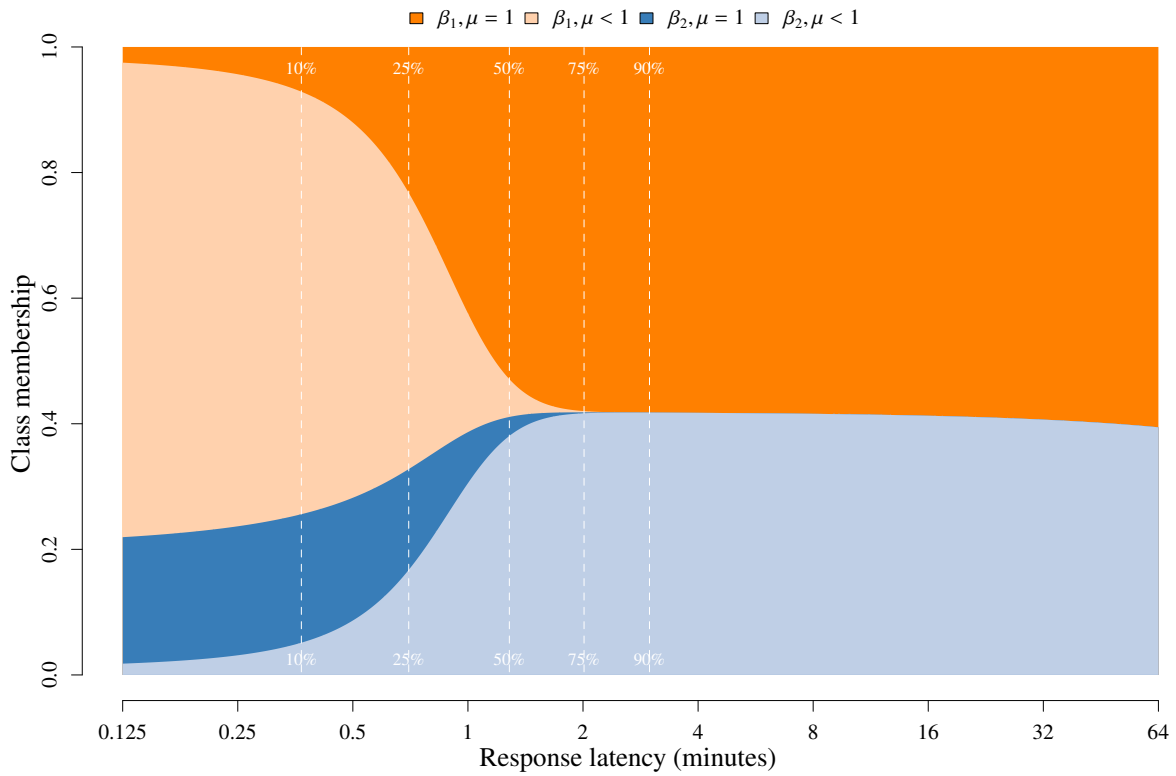


(c) Response latency versus average expected revenue (€ per day)

Figure 3: Response latency versus demand analysis



(a) Response latency versus (unconditional) class membership probabilities for Prb1



(b) Response latency versus (unconditional) class membership probabilities for Prb2

Figure A1: Response latency versus class membership probabilities