# Learning, fatigue and preference formation in discrete choice experiments

Accepted Manuscript


Title: Learning, fatigue and preference formation in discrete choice experiments

Author: Danny Campbell Marco Boeri Edel Doherty W. George Hutchinson

PII: S0167-2681(15)00234-6
DOI: http://dx.doi.org/doi:10.1016/j.jebo.2015.08.018
Reference: JEBO 3657

To appear in: *Journal of Economic Behavior & Organization*

Received date: 19-12-2014
Revised date: 28-8-2015
Accepted date: 29-8-2015

Please cite this article as: Danny Campbell, Marco Boeri, Edel Doherty, W. George Hutchinson, Learning, fatigue and preference formation in discrete choice experiments, *Journal of Economic Behavior and Organization* (2015), http://dx.doi.org/10.1016/j.jebo.2015.08.018

# Learning, fatigue and preference formation in discrete choice experiments

Danny Campbell
Economics Division, Stirling Management School, University of Stirling

Marco Boeri
Gibson Institute, Institute for Global Security, School of Biological Sciences, Queen's University Belfast

Edel Doherty
Department of Economics, University of Galway

W. George Hutchinson
Gibson Institute, Institute for Global Security, School of Biological Sciences, Queen's University Belfast

**Abstract**

While the repeated nature of Discrete Choice Experiments is advantageous from a sampling efficiency perspective, patterns of choice may differ across the tasks, due, in part, to learning and fatigue. Using probabilistic decision process models, we find in a field study that learning and fatigue behavior may only be exhibited by a small subset of respondents. Most respondents in our sample show preference and variance stability consistent with rational pre-existent and well formed preferences. Nearly all of the remainder exhibit both learning and fatigue effects. An important aspect of our approach is that it enables learning and fatigue effects to be explored, even though they were not envisaged during survey design or data collection.

Highlights

- This paper explores learning, fatigue and preference formation in discrete choice experiments.
- The models account for inconsistent preferences and variances at different stages.
- Most respondents in our sample show stability, consistent with rational pre-existent preferences.
- Allowing for learning and fatigue has implications on marginal WTP estimates.

Keywords: discrete choice experiments learning and fatigue behavior preference formation probabilistic decision process model preference and variance consistency

JEL codes: C25 Q51

## 1. Introduction

Discrete Choice Experiments (DCEs) are a stated preference elicitation method, whereby respondents choose their preferred alternative among several hypothetical alternatives in a choice task (e.g., see Louviere et al. (2003) and Hensher et al. (2005) for introductions to the method). The method is widely used for valuing environmental goods and services. In this study we explore preferences for preservation of several rare and endangered fish species in the Lough Melvin Catchment in Ireland using a DCE. As is common practice in DCEs, respondents were asked to consider a number of multidimensional alternatives and to identify their preferred alternative in a choice scenario (or task) where, in our case, different fish species were or were not protected. As in any DCE, in addition to the number of attributes and alternatives per choice task, we had the opportunity to assign the number of choice tasks. In an attempt to increase sampling efficiency we included a large number of choice tasks giving rise to a panel of repeated choice tasks to be completed by each respondent.

A key advantage of using repeated valuation tasks is that they enable researchers to identify the extent to which respondents have clearly defined and established pre-existent preferences for the goods under consideration and the extent to which preferences are modified or even formed through the course of the elicitation process. Despite this, the issues and concerns relating to learning and fatigue are not routinely explored by researchers engaged in stated discrete choice analysis. In this paper we contribute to the literature by proposing a more flexible means for dealing with learning and fatigue in stated preference and more specifically in DCEs conducted in the field. In particular, we explore the extent to which respondents possess or form consistent preferences at different phases in the experiment and whether there is different variability of choice through identification of different scale parameters for each phase.

Our modeling approach builds on the standard multinomial logit (MNL) and random parameters logit (RPL) models, but, unlike previous studies, which have deterministically assumed that the same patterns of learning, fatigue or preference heterogeneity applies to the whole sample, we accommodate the fact that the patterns may be different across respondents. To achieve this we use a probabilistic decision process (PDP) model (e.g. Campbell et al., 2012; McNair et al., 2012; Hensher et al., 2013). This is similar in form to a latent class (LC) model, but the classes here are meant to describe a specific learning and fatigue behavior. The LC model is hence a tool to facilitate differences in learning and fatigue behavior across respondents. As a further departure from the standard LC specification, similar to Greene and Hensher (2013), we facilitate within class random taste variation to capture another layer of preference heterogeneity. We first use this approach to probabilistically determine the proportion of respondents who have consistent preferences as well as preferences that change due to learning or fatigue (or a mixture of the two). We then include scale-adjusted classes, as implemented in Magidson and Vermunt (2008) and Campbell et al. (2011), to ascertain probabilistically the share of respondents with a consistent error variance as well as those who's error variance is different (relative to the middle phase) in the early and/or late phases of the experiment. While both of these PDP models represent an improvement over the existing approaches, they both look at preference and variance consistency in isolation. To overcome this potential weakness, we propose an even more elaborate scale-adjusted PDP model that is aimed at uncovering both types of inconsistency simultaneously. The beneficial feature of this is that we can better disentangle the influence of learning and fatigue upon both the preference parameters as well as the scale parameter. Moreover, it offers a practical approach for DCE practitioners to investigate learning and fatigue, even though they were not considered during survey design or data collection.

Our results show that both learning and fatigue effects are present in this dataset. Our modeling results suggest that, while only a minority of respondents exhibit learning and/or fatigue behavior, expressions of utility (in terms of both preferences and variance) are different in the early and late phases of the experiment (relative to the middle phase) for those respondents identified as exhibiting signs of these patterns. Moreover, our final scale-adjusted PDP specification highlights the potential confound between the two types of inconsistency and, thus, the necessity for specifications that can accommodate both inconsistent preferences and error variance. Results from this model suggest that around two-thirds of respondents have consistent preferences and error variance across the sequence of choice tasks. The remaining respondents are shown to either adjust their preferences or choice variability in approximately equal proportions. Our results also show that model fit as well as marginal willingness to pay (WTP) are impacted by explicitly accommodating learning and fatigue effects on preferences and variability into our models. Our modeling approach also allows us to identify empirically the patterns of responses that may be exhibited in a repeat response DCE as outlined in Day et al. (2012, Table 1, p75). This application of the PDP model can be applied to field datasets to test for patterns associated with "standard" and "non-standard" preference formation. We find that in a large field dataset only a minority of respondents exhibit preference and variance instability but that patterns similar to those identified in Day et al. (2012) can be found. We find that two-thirds of respondents in our study appear to have *a-priori* well formed preferences in terms of demonstrating both preference and variance stability throughout the valuation sequence. One-third show instability of preference and scale throughout the sequence and appear to exhibit preference discovery between the early and middle phase or fatigue between the middle and late phase of the sequence or both. Empirical evidence from our findings suggests that the dominant form of preference and scale instability among this subset of respondents was the combination of preference learning in the early phase of the sequence combined with evidence of fatigue in the late phase.

The remainder of the paper is organized as follows. In the following section, we outline some background to learning and fatigue from a stated preference perspective. In Section 3 we detail our econometric approach and introduce our PDP model with random parameters specification to segment respondents based on their patterns of learning and/or fatigue. In Section 4 we briefly discuss the empirical case-study used to provide data for our analysis. Section 5 reports estimation and post-estimation results while, Section 6 discusses the implications of these findings and concludes.

## 2. Background

There is a well known theoretical and empirical literature suggesting that individuals may exhibit at least two forms of heterogeneity within the sequence of their choices. One type of heterogeneity has been attributed to engaging in some form of learning or discovery process when asked to identify preferences for a sequence of economic goods (see Bradley and Daly, 1994). One of the leading proponents of this learning effect within behavioral economics is Plott (1996), who coined the term "the discovered preference hypothesis". According to Plott, stable and theoretically consistent preferences are formed due to experience gained through practice and repetition and are not necessarily inherent within a decision-maker's initial choices. Plott and Zeiler (2005) demonstrate in a series of economic experiments how major preference anomalies present in initial one-shot valuations such as the WTP willingness to accept (WTA) gap and the "Endowment Effect" are attenuated by repetitive learning which takes place when a subject makes a series of repeat valuations. Within a valuation context two forms of learning have been identified: 'institutional' learning whereby individuals learn the rules of the market (real or hypothetical) and 'value learning' whereby individuals gain knowledge of their own preferences for the good under consideration (Braga and Starmer, 2005). Moreover, both forms of learning are likely to occur when respondents are answering stated preference questions for complex environmental goods, as both the good and market structure are most likely unfamiliar to the individual (especially in the case of non-use value of environmental goods). A second form of heterogeneity that is observed to arise over a sequence of choices is related to the fact that asking respondents to make a large number of complex choices runs the risk that they become fatigued or bored and increasingly confused (Alberini, 2012). In this case it is plausible that their late choices will reflect a higher dimension of variability, as discussed in Hess et al. (2012). Bateman et al. (2008), is the first paper in stated preference to provide evidence for the positive impact of repetition. They observed that within a stated preference study well documented anomalies of anchoring and inconsistency in double-bounded dichotomous choice contingent valuation are attenuated by repetition of valuation task, which they called Learning Design Contingent Valuation (LDCV). DCEs by their design provide a sequence of valuation tasks similar to LDCV. DeShazo and Fermo (2002); Hu et al. (2006); Scarpa et al. (2007); Hess et al. (2012); Carlsson et al. (2012); Czajkowski et al. (2014) all draw attention to potential preference inconsistency that can arise as respondents move through the sequence of choices in DCE.

The issue of preference discovery and the formation of rational consistent and well formed preferences through such a process has been explored in the context of contingent valuation studies (e.g. Bateman et al., 2008; Brouwer, 2012). In DCEs, Swait and Adamowicz (2001); Holmes and Boyle (2005); Scarpa et al. (2007); Savage and Waldman (2008); Day et al. (2012) have explored issues of preference consistency and formation within the sequence of valuations. The primary focus of these studies has been the comparison of results attained under the early valuation tasks against those from the late tasks. The hypothesis is that consistent rational preferences would demonstrate consistency between the early, middle and late stages of the valuation sequence. This very fundamental value consistency test between the value of identical goods at the early, middle and late stages of the valuation is the opposite of the basic scope coherent value tests used in Contingent Valuation scope tests and in the behavioral economics experiments of Ariely et al. (2003), who demonstrate appropriate variations in WTP to changes in the quantity and quality of goods valued.

Another common method for exploring learning and fatigue effects on consistency is to examine the differences in error variance between the early and late choices (Bradley and Daly, 1994; Sælensminde, 2001; Kingsley and Brown, 2010; Oppewal et al., 2010). Lagerkvist et al. (2006) uses a binary heteroskedastic model to account for differences in error variance between different subsets of choice tasks. They find higher variance associated with the late choice tasks, which they attribute to respondent fatigue. On the other hand, Holmes and Boyle (2005) and Kingsley and Brown (2010) find the lowest variance associated with the last choices which they suggest indicates learning effects. In contrast, Carlsson and Martinsson (2001) find no evidence of learning or fatigue effects when they compare results from the same choice tasks given to respondents in a split sample in early and late choice tasks. Brouwer et al. (2010) compare results from self reported certainty and econometric testing procedures, and observe that the self reported statements indicate learning effects, however, the econometric results did not identify any significant differences in parameter estimates or variance across the choice tasks. Day and Prades (2010) also investigated this heterogeneity within the sequence of choices by staging several design treatments to study ordering effects in hypothetical choice surveys. In a follow-up paper, Day et al. (2012) test the strength of ordering effects in sequential choices under stepwise and advanced information treatments introduced into the design and administration. They separate the effects arising into "position dependence" and "precedent dependence" (see Day et al. (2012, Table 1, p75)), which relate to choice task position effects and the influence of previous choice tasks respectively.

## 3. Econometric approach

### 3.1. Background notation

DCEs are an application of the theory of value (Lancaster, 1966), combined with Random Utility Maximization theory (Thurstone, 1927; Manski, 1977). A central assumption of DCEs is that the choice is driven by the maximization of respondents' utility. The utility, denoted by $U$, that each alternative brings to the respondents can be represented by the function:

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

where $n$ indicates the respondent, $i$ the chosen alternative in choice occasion $t$, $\beta$ is a vector of parameters to be estimated for the vector of DCE attributes $x$, and $\varepsilon$ is a random error term (which is unobserved by the researcher) assumed to be an *iid* type I extreme value (EV1) distributed error term, with variance equal to $\pi^2/6\lambda^2$, where $\lambda$ 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:

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

where $y_n$ gives the sequence of choices over the $T_n$ choice occasions for respondent $n$, i.e., $y_n = \langle i_{n1}, i_{n2}, \ldots, i_{nT_n} \rangle$ and $V_{nit} = \beta' x_{nit}$. However, the scale factor, $\lambda$, is typically unidentifiable due to confounding with the vector of parameters $\beta$. For this reason it is usually arbitrary set to one, leading to a constant variance equal to $\pi^2/6$.

While this specification directly uncovers estimates of preferences for each of the DCE attributes, it does so in a manner that assumes that all respondents share the same preferences. While this assumption may hold in some cases, for a variety of reasons one may postulate the hypothesis that it is more likely that the preferences will be heterogeneous across respondents (e.g., see Hensher and Greene (2003), for a discussion). Consequently, in this paper we accommodate this preference heterogeneity via a random parameters logit (RPL) model. Denoting the joint density of $[\beta_{n1}, \beta_{n2}, \ldots, \beta_{nK}]$ by $f(\theta_n|\Omega)$, where $\theta_n$ represents the vector comprised of the random parameters and $\Omega$ denotes the parameters of these distributions (e.g., the mean and variance), the unconditional choice probability under a RPL model is the integral of the logit formula over all possible values of $\beta_{n1}, \beta_{n2}, \ \ldots, \beta_{nK}$:

$$\Pr(y_n|x_n, \Omega) = \int \prod_{t=1}^{T_n} \frac{\exp(\lambda V_{nit})}{\sum_{j=1}^{J} \exp\left(\lambda V_{njt}\right)} f(\theta_n|\Omega)\, d(\theta_n). \tag{3}$$

In this RPL specification parameters of the continuous distributions (i.e., $\Omega$) are obtained. This generally leads to significant gains in model performance and, importantly, greater insights into choice behaviors and preferences for the DCE attributes.

### 3.2. Accommodating learning and fatigue

In this paper we are interested in exploring the effects of learning (as the respondent progresses through the experiment) and fatigue (as the respondent is faced with a long series of choice tasks). Like Czajkowski et al. (2014), our focus is to assess whether the consequences of learning and fatigue lead to inconsistent preferences and error variance across the sequence of choices.

To examine these issues, we introduce three dummy variables: $d_{\mathrm{E}} = 1$ if the choice task is in the early (E) phase (potentially characterized by learning); $d_{\mathrm{M}} = 1$ for the choices in the middle (M) phase; and, $d_{\mathrm{L}} = 1$ for choices in the late (L) phase (potentially characterized by fatigue). In the first instance, it is possible to consider that the impact of learning and fatigue leads to different preferences across the three phases (early, middle and late). In this case the choice probabilities can be described using the following:

$$\Pr\left(y_n|x_n,\Omega\right) = \int \prod_{t=1}^{T_n} \frac{\exp\left(d_{\mathrm{E}}\Delta V_{\mathrm{EM}_{nit}} + d_{\mathrm{M}} V_{\mathrm{M}_{nit}} + d_{\mathrm{L}}\Delta V_{\mathrm{LM}_{nit}}\right)}{\sum_{j=1}^{J} \exp\left(d_{\mathrm{E}}\Delta V_{\mathrm{EM}_{njt}} + d_{\mathrm{M}} V_{\mathrm{M}_{nj}} + d_{\mathrm{L}}\Delta V_{\mathrm{LM}_{njt}}\right)} f\left(\theta_n|\Omega\right) d\left(\theta_n\right), \tag{4}$$

where $V_{\mathrm{M}_{nit}}$ represents the observed utility function for the group of choice tasks in the middle of the experiment, while $\Delta V_{\mathrm{EM}_{nit}}$ and $\Delta V_{\mathrm{LM}_{nit}}$ respectively denote the utility functions at early and late phases of the experiment, expressed in differences from the middle phase. This eases interpretation and enables at a glance an assessment of whether or not the parameters in the early and late phases are significantly different from the parameter estimates from the middle phase, which is our main interest.

A first step in many studies investigating learning and fatigue is an assessment of differences in variance throughout the sequence. We, therefore, continue with this line of inquiry. To do so we allow for scale differences, but fix preferences, across the three phases. We achieve this by estimating different scale parameters for both the early and late phases. As a baseline, we set the scale parameter for the middle set of choices to unity (i.e., $\lambda_{\mathrm{M}} = 1$), as follows:

$$\Pr\left(y_n|x_n,\Omega\right) = \int \prod_{t=1}^{T_n} \frac{\exp\left(\left(\lambda_{\mathrm{M}} + d_{\mathrm{E}}\eta_{\mathrm{E}} + d_{\mathrm{L}}\eta_{\mathrm{L}}\right)\beta_n' x_{nit}\right)}{\sum_{j=1}^{J} \exp\left(\left(\lambda_{\mathrm{M}} + d_{\mathrm{E}}\eta_{\mathrm{E}} + d_{\mathrm{L}}\eta_{\mathrm{L}}\right)\beta_n' x_{njt}\right)} f\left(\theta_n|\Omega\right) d\left(\theta_n\right), \tag{5}$$

where $\eta_{\mathrm{E}}$ and $\eta_{\mathrm{L}}$ are defined as differences from $\lambda_{\mathrm{M}}$ (i.e., $\eta_{\mathrm{E}} = \lambda_{\mathrm{E}} - \lambda_{\mathrm{M}}$ and $\eta_{\mathrm{L}} = \lambda_{\mathrm{L}} - \lambda_{\mathrm{M}}$ respectively). Again, this is intended to facilitate the comparison against the middle choices, which is our central interest.

While equations [4] and [5] allow for the possibility of inconsistent preferences and variance respectively, there is a risk that they do not fully describe the changes that occur during the three phases of the experiment. This stems from the fact that both specifications are based on the underlying assumption that all respondents follow the same pattern of preference changes in the former and variance changes in the latter. However, this is likely to be an unrealistic assumption, since the exhibited patterns of learning and fatigue behavior may vary across respondents. For this reason, we exploit the LC framework approach to better accommodate these patterns. While based on a LC model specification, we prefer to describe our models as probabilistic decision process (PDP) models (Campbell et al., 2012, 2014), since each latent class is described by a specific decision process that is linked with learning and fatigue rather than marginal utilities (e.g., see Erdem et al., 2014, 2015; Campbell and Erdem, 2015, for further recent applications of PDP models to explore a range of behavioral heuristics adopted by respondents during stated preference studies). We note that we continue with the RPL framework, which is also a departure from the standard LC specification (i.e., similar to those presented in Greene and Hensher (2013) and Campbell et al. (2014), the models capture another layer of preference heterogeneity within each class).

In terms of studying inconsistent preferences in responses to learning and fatigue at least four distinct subgroups (i.e., classes) can be identified: (i) those who do not exhibit any learning or fatigue behavior and, thus, have consistent preferences across the entire experiment; (ii) those who exhibit learning and have a different preference structure during the early phase relative to the middle and late phases; (iii) those who exhibit fatigue and have a different preference structure during the late phase; and, finally (iv) those who exhibit learning and fatigue and have different preference structures during the early, middle and late phases. To uncover these four subgroups (denoted by $g$), similar to the approach implemented in Scarpa et al. (2009), the estimated utility coefficients are restricted to take the same value across all four classes, but where specific parameters based on the early choice tasks are obtained only in classes 2 and 4 and parameters derived from the late choice tasks are uncovered only in classes 3 and 4, described as follows:

$$\begin{aligned}
V_{g_{1}nit} &= V_{\mathrm{M}_{nit}} \\
V_{g_{2}nit} &= d_{\mathrm{E}}\Delta V_{\mathrm{EM}_{nit}} + (d_{\mathrm{M}} + d_{\mathrm{L}})\, V_{\mathrm{M}_{nit}} \\
V_{g_{3}nit} &= (d_{\mathrm{E}} + d_{\mathrm{M}})\, V_{\mathrm{M}_{nit}} + d_{\mathrm{L}}\Delta V_{\mathrm{LM}_{nit}} \\
V_{g_{4}nit} &= d_{\mathrm{L}}\Delta V_{\mathrm{EM}_{nit}} + d_{\mathrm{M}} V_{\mathrm{M}_{nit}} + d_{\mathrm{E}}\Delta V_{\mathrm{LM}_{nit}}.
\end{aligned} \tag{6a}$$

The first class, $V_{g_1}$, in equation [6a] specifies that preferences are consistent throughout the sequence of choice tasks (i.e., $\Delta V_{\mathrm{EM}} = \Delta V_{\mathrm{LM}} = 0$), which is equivalent to equation [3]. The only difference in the second class, $V_{g_2}$, is that $\Delta V_{\mathrm{EM}} \neq 0$, meaning that differences in preferences can be recovered from choices made in the early phase. Compared to the first class, the third class, $V_{g_3}$, specifies that $\Delta V_{\mathrm{LM}} \neq 0$, which allows a different set of preference coefficients to be retrieved from the late phase. The final class, $V_{g_4}$, allows different preferences across the three phases (i.e., $\Delta V_{\mathrm{EM}} \neq 0$ and $\Delta V_{\mathrm{LM}} \neq 0$), and is tantamount to equation [4].

A respondent's true learning and fatigue behavior cannot be known with certainty and, thus, remains latent. To work around this, based on observed choice behavior, probabilistic conditions can be imposed on the utility expressions in equation [6a]. In doing so, the presence of each expression can be established up to a probability, with the full probability per respondent allocated across all $G$ classes. Under such a PDP model it is possible to estimate the probabilities of each subgroup, with the overall probability of choice given by:

$$\Pr(y_n|x_n,\Omega) = \sum_{g=1}^{4} \pi_g \int \prod_{t=1}^{T_n} \frac{\exp\left(V_{g_{nit}}\right)}{\sum_{j=1}^{J} \exp\left(V_{g_{njt}}\right)} f\left(\theta_{n_{g_1}}|\Omega_{g_1}\right) d\left(\theta_{n_{g_1}}\right), \tag{6b}$$

where $\pi_g$ denotes the unconditional probability associated with observing the utility function relating to class $g$ (i.e., the prior likelihood of competing learning and fatigue behaviors being their actual behavior).

In a similar vein, it may be appropriate to allow for variance differences instead of preference differences across the three phases of the experiment. Similar to the segmentation used above, when considering the possible variance differences respondents can be considered as belonging to one of four subgroups: (i) those who do not exhibit any learning or fatigue behavior and, thus, have a consistent variance across the entire experiment; (ii) those who exhibit learning and have a different variance during the early phase of the experiment; (iii) those who exhibit fatigue and have a different variance for the late choices; and, finally (iv) those who exhibit learning and fatigue and have a different variance for the early and late phases compared to the middle phase. Building on the notation of equation [5], the scale parameters for each subgroup can be described as follows:

$$\begin{aligned}
\mu_{s_1} &= \lambda_{\mathrm{M}} \\
\mu_{s_2} &= d_{\mathrm{E}}\eta_{\mathrm{E}} + (d_{\mathrm{M}} + d_{\mathrm{L}})\,\lambda_{\mathrm{M}} \\
\mu_{s_3} &= (d_{\mathrm{E}} + d_{\mathrm{M}})\,\lambda_{\mathrm{M}} + d_{\mathrm{L}}\eta_{\mathrm{L}} \\
\mu_{s_4} &= d_{\mathrm{E}}\eta_{\mathrm{E}} + d_{\mathrm{M}}\lambda_{\mathrm{M}} + d_{\mathrm{L}}\eta_{\mathrm{L}}.
\end{aligned} \tag{7a}$$

The first class, $\mu_{s_1}$, in equation [7a] assumes that error variance is stable over the entire experiment (i.e., $\eta_{\mathrm{E}} = \eta_{\mathrm{L}} = 0$), which is also analogous to equation [3]. The second class, $\mu_{s_2}$, deviates from the first class in that $\eta_{\mathrm{E}} \neq 0$ so that a different scale parameter is obtained for the early phase. In contrast, the third class, $\mu_{s_3}$, assumes $\eta_{\mathrm{L}} \neq 0$, thereby capturing any difference in the variance during the late phase. The final class, $\mu_{s_4}$, allows different scale parameters in all three phases (i.e., $\eta_{\mathrm{E}} \neq 0$ and $\eta_{\mathrm{L}} \neq 0$), and is comparable to equation [5].

Therefore, the choice probability can be described by a scale-adjusted PDP (cf., Magidson and Vermunt, 2008; Campbell et al., 2011) model as follows:

$$\Pr\left(y_n | x_n, \Omega\right) = \sum_{s=1}^{4} \pi_s \int \prod_{t=1}^{T_n} \frac{\exp\left(\mu_s V_{nit}\right)}{\sum_{j=1}^{J} \exp\left(\mu_s V_{njt}\right)} f\left(\theta_n | \Omega\right) d\left(\theta_n\right), \tag{7b}$$

where $\pi_s$ is the unconditional probability relating to $\mu_s$.

Despite the advantages of using the flexible specifications outlined in equations [6] and [7], they both have a weakness since they explore the issues in isolation (i.e., they have the strict assumption of a consistent variance and consistent preferences across all choice tasks respectively). Nevertheless, it is also conceivable that there may be some cases where learning and fatigue manifests itself as inconsistent preferences for some respondents and an inconsistent error variance for others, which leads to seven possible situations outlined below:

$$\begin{aligned}
&W_{q_1 nit} = \mu_{s_1} V_{g_1 nit} && \text{Consistent preferences and consistent variance} \\
&\left.\begin{aligned} W_{q_2 nit} &= \mu_{s_1} V_{g_2 nit} \\ W_{q_3 nit} &= \mu_{s_1} V_{g_3 nit} \\ W_{q_4 nit} &= \mu_{s_1} V_{g_4 nit} \end{aligned}\right\} && \text{Inconsistent preferences and consistent variance} \\
&\left.\begin{aligned} W_{q_5 nit} &= \mu_{s_2} V_{g_1 nit} \\ W_{q_6 nit} &= \mu_{s_3} V_{g_1 nit} \\ W_{q_7 nit} &= \mu_{s_4} V_{g_1 nit} \end{aligned}\right\} && \text{Consistent preferences and inconsistent variance}
\end{aligned} \tag{8a}$$

We, therefore, propose a further PDP model to facilitate these possible changes across the three phases of the experiment:

$$\Pr\left(y_n|x_n,\Omega\right) = \sum_{q=1}^{7}\pi_q \int \prod_{t=1}^{T_n} \frac{\exp\left(W_{q_{nit}}\right)}{\sum_{j=1}^{J}\exp\left(W_{q_{njt}}\right)} f\left(\theta_{n_{g_1}}|\Omega_{g_1}\right) d\left(\theta_{n_{g_1}}\right). \tag{8b}$$

The advantage of this model is that it simultaneously retrieves unconditional probabilities (i.e., $\pi_q$) associated with the four patterns of preferences along with the four patterns of variance that arise from the effects of learning and fatigue.[1]

Incidentally, we note the resemblance of the types of behavior characterized by the classes of our PDP models to the theories of order effects in repeat stated preference studies detailed in Day et al. (2012, Table 1, p75). So, while our PDP models are not intended to classify respondents according to different theories of order effects, we can, nevertheless, produce good insight into whether respondent preferences are pre-existent and well formed or appear to show patterns of learning following the Discovered Preference Theory of Plott and/or of fatigue with increasing variance in later choices as often found in field DCE studies (Hess et al., 2012; Czajkowski et al., 2014). This is an especially important feature of our PDP approach—learning and fatigue behavioral characteristics can be identified even though they were not considered during either survey design or data collection. Our PDP modeling approach does not require the choice tasks to be carefully and systematically arranged (e.g., as in Day et al., 2012) so that the ordering effects can be distinguished. The approach can be applied to any DCE dataset, as long as it is comprised of a panel of repeated choice tasks (which is typically the case).

A unique contribution of this paper to both the preference formation literature and the DCE literature is that our PDP models show that recognized patterns of learning and fatigue differ across major segments of the population. However, it should be noted that the probability estimates ($\pi_g$, $\pi_s$ and $\pi_q$ in equations [6], [7] and [8] respectively) are unconditional and, therefore, do not directly provide any information on the likely learning and fatigue behavior exhibited by a given respondent. Nonetheless, conditional on their choices, and the estimated parameters it is possible to retrieve respondent-specific probabilistic statements about the likelihood of class membership, as follows:

$$\mathbb{E}\left(\hat{\pi}_{q_n}\right) = \frac{\hat{\pi}_q \sum_{q=1}^{Q} \Pr\left(y_n|x_n,\beta_{r_q},\mu_q\right)}{\sum_{q=1}^{Q}\hat{\pi}_q \sum_{r=1}^{R} \Pr\left(y_n|x_n,\beta_{r_q},\mu_q\right)}, \tag{9a}$$

where $\mathbb{E}\left(\hat{\pi}_{q_n}\right)$ represents the expected value of membership to class $q$ for respondent $n$, $\hat{\pi}_q$ is the prior (unconditional) estimate for class $s$, and where $\Pr\left(y_n|x_n,\beta_{r_q},\mu_q\right)$ gives the probability of observing the sequence of choices by respondent $n$ given $x_n$ and the values of $\beta_r$ and $\mu$ relating to class $q$. Here $\beta_r$, with $r = 1,\ldots,R$, represents an independent random draw with equal weight from $f\left(\theta_q|\hat{\Omega}_q\right)$.

Also of key importance is the extent to which the estimates of marginal WTP are sensitive to the manner in which learning and fatigue is accounted for. For this reason, we recover the mean of the conditional marginal WTP distribution for every respondent, $\mathbb{E}\left(\hat{W}_n\right)$, using the following

[1]It is important to recognise that observed changes in variance may be due to preference parameters changing in the same proportion (see Hess and Rose, 2012, for a discussion).

expression:

$$\mathbb{E}\left(\hat{W}_n\right) = \sum_{q=1}^{Q} \mathbb{E}\left(\hat{\pi}_{q_n}\right) \frac{\sum_{r=1}^{R} \Pr\left(y_n | x_n, \beta_{r_q}, \mu_q\right) W_{r_q}}{\sum_{r=1}^{R} \Pr\left(y_n | x_n, \beta_{r_q}, \mu_q\right)}, \tag{9b}$$

where $W_{r_q} = -\beta_{m_{r_q}} / \beta_{\$_{r_q}}$ is the marginal WTP for attribute $m$ (i.e., minus the ratio of the coefficients for attribute $m$ and the cost \$) obtained from the $r^{\text{th}}$ draw from $f\left(\theta_q | \hat{\Omega}_q\right)$.

Note that these conditional parameters themselves follow a distribution, equation [9] merely gives the expected value of these distributions (e.g., see Hess, 2010, for further details). Nevertheless, this does give us some information about the most likely position of a respondent on the distributions of class memberships and marginal WTP, which is generally of greatest interest.

### *3.3. Model estimation*

All models are estimated using Ox (Doornik, 2007). With the exception of the MNL model, the choice probabilities in the models documented above cannot be calculated exactly (because the integrals do not have a closed form). For this reason, we estimate them by simulating the log-likelihood with $R = 300$ quasi-random draws via Halton sampling.[2] In the case of the PDP models (i.e., those that are based on probabilistic class segmentation), we are also mindful of the fact that models of this form are subject to local maxima. In an attempt to reduce the likelihood of reaching a local rather than a global maximum, we employ the BFGS algorithm which has proven good performance even for non-smooth optimizations (Yu et al., 2010) and use a variety of random starting points. Specifically, we do this by estimating the models many times (at least 10 times), but each time using a different vector of starting values, which are chosen randomly.

A key element with the specification of random taste variation is the assumption regarding the distribution of each of the random parameters (Hensher and Greene, 2003; Hess et al., 2005). Random parameters can take a number of predefined functional forms. While this affords the analyst with some control and flexibility, the random parameters are not observed and there is typically little *a priori* information about the shape of its distribution except possibly a sign constraint (Fosgerau and Hess, 2009). Consequently, the chosen distribution is essentially an arbitrary approximation (Hensher and Greene, 2003) requiring some possibly strong or unwarranted distributional assumptions about individual heterogeneity (Greene and Hensher, 2003). After evaluating the results from various specifications and distributional assumptions, we specify all parameters, with the exception of the Cost parameter[3], within the vector $\beta_{\mathrm{M}}$ as having Normal distributions: $\beta_{k,\mathrm{M}} = \mu_{k,\mathrm{M}} + \sigma_{k,\mathrm{M}} \upsilon_k$, where $\upsilon_k$ is an independent standard Normal deviate and $\mu_k$ and $\sigma_k$ are parameters to be estimated, which can be interpreted as the mean and the standard deviation respectively of the $k$th Normally distributed parameter. While we acknowledge that our decision to assume independence between the random parameters is a limitation, we did find that our data was suitably characterized by our choice of distributions

[2] We recognise that 300 draws is relatively low. We justify this on the grounds that we tested 85 random parameter specifications using multiple starting points. Increasing the number of draws would have entailed considerably more estimation time.

[3] After evaluating a number of continuous distributions, we settled on a non-random Cost parameter. We remark that this greatly facilitated calculating the distributions of marginal WTP and it avoided the prediction of extreme WTP outliers, which we feared might bias our marginal WTP comparisons. We also note that our decision to fix the Cost parameter should not preclude us from exploring the issue which is at hand, namely the incidence of learning and/or fatigue behavior.

given consideration to the plausible signs on the coefficients and with regard to the evaluation of the log-likelihood values using different distributions.

## 4. Empirical case-study

To illustrate the proposed methodology on an empirical case-study we use stated preference data collected to estimate the existence value of rare and endangered fish species in the Lough Melvin Catchment in Ireland. Lough Melvin is a freshwater lake in the North West of Ireland which straddles the border between Northern Ireland and the Republic of Ireland. With a unique population of native fish species, the Lough Melvin Catchment has an internationally important conservation status. Lough Melvin and its associated river system supports the only remaining population of Arctic char *Salvelinus alpinus* (L.) in Northern Ireland and contains Atlantic salmon *Salmo salar* (L.) and three genetically distinct populations of brown trout known as ferox *Salmo ferox* (L.), gillaroo *Salmo stomachicus* (L.) and sonaghan *Salmo nigripinnis* (L.). Since the habitat of these fish populations is recognized as being vulnerable, there was a need to assess the extent to which the general public supports the prevention of their extinction.

The DCE consisted of a panel of sixteen repeated choice tasks. To control for anchoring or focalism a number of different versions were used, each of which had a different sequence of the choice tasks. While respondents were informed that they would answer a sequence of choice tasks, they were not informed how many. Each choice task outlined three possible outcomes. The first two outcomes—labeled as 'Option A' and 'Option B'—described the conservation status of each of the fish species after the implementation of two experimentally designed conservation schemes. At the end of these schemes, the fish species would either be 'Conserved' or 'Extinct'. While a particular scheme described under either 'Option A' or 'Option B' may have been unable to prevent some of the fish species from becoming extinct, they both ensured against the extinction of all fish species (i.e., at least one species was conserved under each scheme). The final outcome—labeled as 'Do Nothing'—showed the expected outcome if nothing was done to protect the fish species. In this case, the respondents were informed that all five fish species would become extinct. 'Option A' and 'Option B' were both described to respondents as available at a positive cost. The payment vehicle used was the amount that they would personally have to pay—through a one-off increase in their Income Tax and/or Value Added Tax contributions—to implement the scheme. The 'Do Nothing' (or status quo) option had zero cost to the respondent.

The DCE exercise reported here involved several rounds of design and testing to ensure descriptions of the attribute levels were meaningful. The process began with a qualitative review of opinions from those involved in the design and implementation of the Lough Melvin Catchment Management Plan. Further qualitative research was carried out to refine the definitions of the attribute levels so that they could be used in the survey. This was achieved through a series of focus group discussions with members of the general public. In total, six focus group discussions were held: two within the Lough Melvin Catchment area; and, four in other areas throughout Northern Ireland and the Republic of Ireland. Participants comprised of several key user groups, including anglers, farmers, foresters and members of the general public. The groups typically included between eight to ten participants (from the same or similar user group). The discussions generally lasted approximately one hour and were structured around a general discussion of the relevant issues and testing of the face validity of the questionnaire. Feedback from this helped to refine the comprehension, options, layout and wording of the DCE. Pilot testing, involving over 100 respondents, of the survey instrument was also conducted in the field. This allowed the collection of additional information, which along with expert judgment and observations from the focus group discussions, was used to identify and refine the attributes. Feedback gathered during this process and the pilot study revealed that people did find the experiment meaningful and credible.

The population of interest was the adult population of the Republic of Ireland and Northern Ireland. The study adopted a stratified random sample to reflect the geographic distribution of the adult population; the approximate rural/urban split; the approximate socio-economic status of the regional population; and the approximate gender and age profile of the populations within both jurisdictions. A final sample of 624 usable interviews was obtained which, with each respondent answering 16 choice tasks, resulted in 9,984 observations for model estimation. For further analysis on this data interested readers are referred to Campbell et al. (2010) and Campbell et al. (2012).

## 5. Results

As part of our analysis, rather than use an arbitrary definition of early, middle and late choices we ran a series of models with different number of choice tasks in each of the three phases. Conducting this type of investigation also gives an insight into the most likely patterns of learning and fatigue (i.e., when learning is most likely to end and fatigue is most likely to begin) for a given dataset. Based on this investigation we observe that the most significant gains in model fit (for our empirical case-study) are generally attained when we define the first eight choice tasks as early choices, with the remaining eight choice tasks split equally between middle and late choices. For this reason, we focus on this pattern in the remaining part of the paper. Nevertheless, we provide summary results for the remaining patterns in Appendix A.

This split into the early, middle and late phases for choices has similarities to Savage and Waldman (2008) and Carlsson et al. (2012) who split their panel of choices into a first half and last half sequence of choices. Carlsson et al. (2012) also analyze the first choice separately, as did Hess et al. (2012) and Czajkowski et al. (2014) both of whom analyses each choice in the panel separately. These papers all found that the first choice in the panel produced different value estimates and scale parameters to subsequent choices as was also found very clearly for repeated choices using contingent valuation by Bateman et al. (2008). While the first time may be "the hardest" (Carlsson et al., 2012) the considerable length of panels and methods of analysis used in DCEs may minimize this effect to reveal a major class of respondents with consistent preferences and scale parameters throughout the sequence.

### *5.1. Estimation results*

Table 1 reports estimation results obtained from six model specifications. Model 1 relates to the RPL model (equation [3]), with a marginal utility parameter for the cost attribute and a mean and standard deviation for a fish attribute and status quo alternative specific constant, which can be interpreted as the distribution parameters describing the marginal (dis)utility from the extinction of *all* fish species (whereas the parameters associated with the fish attribute describes the marginal (dis)utility distribution associated with preserving *one* species).[4] In line with *a-priori* expectations, the cost coefficient is estimated as having a negative, and significant, sign—implying that respondents, *ceteris paribus*, prefer policy scenarios that are less expensive. As

[4]The fish attribute included in the model represents the number of rare and endangered species protected under each policy scenario. Although this is quite restrictive, since it does not allow for non-linear effects to be tested nor differentiate between the species, it has the advantage of being parsimonious and sufficient for the purpose at hand, which is the assessment of learning and fatigue behavior.

Table 1: Estimation results

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| $\Delta\beta_{\text{Cost,EM}}$ | | -0.002 (0.002) | | -0.249$^{***}$ (0.031) | | -0.379$^{***}$ (0.043) |
| $\beta_{\text{Cost,M}}$ | -0.019$^{***}$ (0.001) | -0.019$^{***}$ (0.002) | -0.019$^{***}$ (0.001) | -0.009$^{***}$ (0.001) | -0.017$^{***}$ (0.001) | -0.010$^{***}$ (0.001) |
| $\Delta\beta_{\text{Cost,LM}}$ | | 0.005$^{*}$ (0.003) | | -0.436$^{***}$ (0.078) | | -0.619$^{***}$ (0.108) |
| $\Delta\beta_{\text{Fish,EM}}$ | | 0.034 (0.040) | | 0.380$^{***}$ (0.119) | | 0.180 (0.129) |
| $\mu_{\text{Fish,M}}$ | 0.486$^{***}$ (0.033) | 0.495$^{***}$ (0.043) | 0.483$^{***}$ (0.038) | 0.477$^{***}$ (0.031) | 0.382$^{***}$ (0.015) | 0.419$^{***}$ (0.022) |
| $\sigma_{\text{Fish,M}}$ | 0.684$^{***}$ (0.034) | -0.698$^{***}$ (0.033) | 0.677$^{***}$ (0.043) | 0.576$^{***}$ (0.033) | 0.529$^{***}$ (0.020) | 0.370$^{***}$ (0.021) |
| $\Delta\beta_{\text{Fish,LM}}$ | | -0.094$^{**}$ (0.044) | | -0.092 (0.196) | | -0.580$^{**}$ (0.233) |
| $\Delta\beta_{\text{SQ,EM}}$ | | -0.193 (0.144) | | -2.347$^{***}$ (0.401) | | -3.699$^{***}$ (0.484) |
| $\mu_{\text{SQ,M}}$ | -3.219$^{***}$ (0.201) | -3.079$^{***}$ (0.205) | -3.183$^{***}$ (0.237) | -3.638$^{***}$ (0.254) | -2.714$^{***}$ (0.105) | -3.822$^{***}$ (0.209) |
| $\sigma_{\text{SQ,M}}$ | 2.419$^{***}$ (0.185) | -2.387$^{***}$ (0.150) | 2.371$^{***}$ (0.214) | 2.955$^{***}$ (0.259) | 2.285$^{***}$ (0.087) | -3.252$^{***}$ (0.173) |
| $\Delta\beta_{\text{SQ,LM}}$ | | 0.003 (0.159) | | -4.950$^{***}$ (1.019) | | -7.089$^{***}$ (1.324) |
| $\eta_{\text{E}}$ | | | 0.097 (0.061) | | 46.527$^{***}$ (11.112) | 15.258$^{***}$ (3.907) |
| $\eta_{\text{L}}$ | | | -0.122$^{**}$ (0.056) | | 32.987$^{**}$ (14.909) | 48.920 (57.509) |
| $\pi_{q_1}$ | 1.000 | | | 0.781$^{***}$ (0.038) | 0.777$^{***}$ (0.044) | 0.659$^{***}$ (0.042) |
| $\pi_{q_2}$ | | | | 0.073$^{***}$ (0.015) | | 0.045$^{***}$ (0.011) |
| $\pi_{q_3}$ | | | | 0.010 (0.007) | | 0.010 (0.007) |
| $\pi_{q_4}$ | | 1.000 | | 0.135$^{***}$ (0.017) | | 0.111$^{***}$ (0.015) |
| $\pi_{q_5}$ | | | | | 0.076$^{***}$ (0.018) | 0.007 (0.022) |
| $\pi_{q_6}$ | | | | | 0.000 (0.022) | 0.094$^{***}$ (0.018) |
| $\pi_{q_7}$ | | | 1.000 | | 0.147$^{***}$ (0.020) | 0.075$^{***}$ (0.019) |
| $\mathcal{L}(\hat{\beta})$ | -7,380.858 | -7,367.122 | -7,372.373 | -7,034.805 | -7,230.828 | -6,992.752 |
| $K$ | 5 | 11 | 7 | 14 | 10 | 19 |
| $\bar{\rho}^2$ | 0.327 | 0.327 | 0.327 | 0.357 | 0.340 | 0.361 |
| AIC | 14,771.716 | 14,756.245 | 14,758.746 | 14,099.611 | 14,483.656 | 14,025.504 |
| BIC | 14,775.925 | 14,754.453 | 14,760.955 | 14,093.820 | 14,481.864 | 14,014.712 |

*Notes:*

1. Standard errors in parentheses ($^{*}$ significant at the 10% level; $^{**}$ significant at the 5% level; and, $^{***}$ significant at the 1% level).
2. The associated null hypothesis for $\pi_{q_1}$ is $H_0$: $\pi_{q_1} = 1$.

anticipated, the mean coefficient of the fish attribute is estimated as having a positive, and statistically significant, sign—suggesting that respondents, all else held constant, prefer the fish species to be conserved (relative to extinction). In connection with this, the status quo mean coefficient is negative, and significant—indicating that, *ceteris paribus*, respondents have a non-linear dislike of all the fish species becoming extinct. Inspecting the standard deviations of the fish attribute and the status quo reveals that they are also significant, which means we can reject the null hypothesis of preference homogeneity.

Model 2 (equation [4]) is aimed at accounting for the fact that the effects of learning and fatigue may manifest as inconsistent preferences across the sequence of choices made by individuals. This model retrieves separate parameters for the early, middle and late phases of the experiment. In particular, these parameters determine the 'location' or shift in the distribution (e.g., if $\Delta\beta > 0$ the probability density or mass function shifts rigidly to the right, maintaining its exact shape).[5] A comparison of the coefficients uncovered for the middle sequence of choices against the respective coefficients attained under Model 1 reveals that somewhat similar inferences can be made. From this model we do not find any significant change in price sensitivity during the sequence of choices. In contrast, there is evidence that the marginal utilities associated with the fish attribute decrease during the late choices. Similar to the cost attribute, we do not find evidence of changes in marginal utilities linked with the status quo. We observe an improvement in model fit over Model 1, with the likelihood ratio test statistic of 27.47 against the $\chi^2$ critical value of 12.59 $\left(\chi^2_{6,0.05}\right)$.

While Model 2 gives some indication of the manner in which respondents' preferences, and potentially the decision-making strategies and heuristics they adopted, change during the experiment, the fact that the negatively and positively estimated parameters move in opposite directions, suggests that scale may also be at play. In Model 3 (equation [5]) we explore a common starting point for assessing the effects of learning and fatigue which stems from the notion that the variance of respondent's choices differ along the experiment. An examination of our findings highlights that the variance is highest during the final sequence of choice tasks (i.e., decreasing scale parameter) While inferences relating to the other coefficients remain unchanged from Model 1, we note that this model is associated with an improved fit. Indeed, the improvement of over 8 log-likelihood units at the expense of two additional parameters provides a likelihood ratio test statistic of 16.97 against the $\chi^2$ critical value of 5.99 $\left(\chi^2_{2,0.05}\right)$. However, in terms of model fit, Model 3 is inferior to Model 2, which gives a signal, for this dataset at least, the greater importance of recognizing inconsistent preferences compared to error variance when studying learning and fatigue.

Models 2 and 3 both offer advantages over Model 1. However, they are both based on a deterministic treatment of learning and fatigue behavior, whereby all respondents are expected to exhibit the same ordering effects. For this reason, we now turn to our PDP specifications.

In Model 4 (equation [6]) we consider a PDP specification to accommodate the different patterns of learning and fatigue behavior set out in equation [6a]. Importantly, the results from this model suggest that learning and/or fatigue behavior (in terms of inconsistent preferences) applies only to a subset of respondents. In fact, the unconditional probability retrieved for utility function $V_{g_1}$ (i.e., the class defined as having consistent preferences) in equation [6a] is almost 80 percent. Indeed, looking the summary statistics of the means of the conditional class proba-

[5]Note that it would also be possible to accommodate for potential differences in the spread of the random parameters during different phases of the DCE. While allowing for this is likely to improve model fit and give a deeper insight into learning and fatigue patterns, keeping the spreads equal has the appeal that the differences in the distributions can be envisaged more straightforwardly.

bility, derived using equation [9a], for this class in Table 2, we see that over half are effectively one. Nevertheless, the fact that the unconditional probability is significantly different from 1 does suggest that the assumption of consistent preferences is inappropriate. The majority of the remaining respondents display signs of learning and fatigue (the unconditional class probability relating to utility function $V_{g_4}$ in equation [6a] is 14 percent). Interestingly, the approximate 20 percent—based on combined unconditional probabilities for utility functions $V_{g_2}$ and $V_{g_4}$, 0.073 and 0.135 respectively—who exhibited signs of learning were significantly more price sensitive in the early phase of the experiment and had significantly higher marginal utilities for conserving the fish species. Connected with this, is the significantly lower estimated value for the status quo constant. This finding may signal different processing strategies of attributes, warm-glow, yeah-saying and other forms of strategic bias as well as using price as a proxy for quality, all of which are likely to be more prominent in the early choices before respondents have established their decision-making mechanism and preferences (i.e., institutional and value

Table 2: Summary of the means of the conditional class probability distributions

| | 5%ile | 10%ile | 25%ile | Median | Mean | 75%ile | 90%ile | 95%ile | % predicted modal class |
|---|---|---|---|---|---|---|---|---|---|
| Model 4 | | | | | | | | | |
| $\mathbb{E}(\hat{\pi}_{g_1})$ | 0.002 | 0.019 | 0.734 | 1.000 | 0.781 | 1.000 | 1.000 | 1.000 | 0.787 |
| $\mathbb{E}(\hat{\pi}_{g_2})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.073 | 0.020 | 0.197 | 0.623 | 0.066 |
| $\mathbb{E}(\hat{\pi}_{g_3})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.010 | 0.000 | 0.006 | 0.026 | 0.005 |
| $\mathbb{E}(\hat{\pi}_{g_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.135 | 0.002 | 0.876 | 0.969 | 0.143 |
| $\mathbb{E}(\hat{\pi}_{g_2}+\hat{\pi}_{g_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.208 | 0.234 | 0.979 | 0.996 | |
| $\mathbb{E}(\hat{\pi}_{g_3}+\hat{\pi}_{g_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.145 | 0.010 | 0.883 | 0.970 | |
| Model 5 | | | | | | | | | |
| $\mathbb{E}(\hat{\pi}_{s_1})$ | 0.054 | 0.141 | 0.476 | 1.000 | 0.777 | 1.000 | 1.000 | 1.000 | 0.744 |
| $\mathbb{E}(\hat{\pi}_{s_2})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.076 | 0.067 | 0.244 | 0.525 | 0.071 |
| $\mathbb{E}(\hat{\pi}_{s_3})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\mathbb{E}(\hat{\pi}_{s_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.147 | 0.009 | 0.726 | 0.862 | 0.186 |
| $\mathbb{E}(\hat{\pi}_{s_2}+\hat{\pi}_{s_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.223 | 0.524 | 0.859 | 0.946 | |
| $\mathbb{E}(\hat{\pi}_{s_3}+\hat{\pi}_{s_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.147 | 0.009 | 0.726 | 0.862 | |
| Model 6 | | | | | | | | | |
| $\mathbb{E}(\hat{\pi}_{q_1})$ | 0.001 | 0.012 | 0.223 | 0.916 | 0.659 | 0.998 | 1.000 | 1.000 | 0.659 |
| $\mathbb{E}(\hat{\pi}_{q_2})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.045 | 0.001 | 0.055 | 0.269 | 0.042 |
| $\mathbb{E}(\hat{\pi}_{q_3})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.010 | 0.000 | 0.003 | 0.046 | 0.003 |
| $\mathbb{E}(\hat{\pi}_{q_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.111 | 0.000 | 0.732 | 0.969 | 0.117 |
| $\mathbb{E}(\hat{\pi}_{q_5})$ | 0.000 | 0.000 | 0.000 | 0.001 | 0.075 | 0.049 | 0.260 | 0.515 | 0.058 |
| $\mathbb{E}(\hat{\pi}_{q_6})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.007 | 0.005 | 0.028 | 0.038 | 0.000 |
| $\mathbb{E}(\hat{\pi}_{q_7})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.094 | 0.017 | 0.540 | 0.654 | 0.122 |
| $\mathbb{E}(\hat{\pi}_{q_2}+\hat{\pi}_{q_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.156 | 0.011 | 0.939 | 0.993 | |
| $\mathbb{E}(\hat{\pi}_{q_3}+\hat{\pi}_{q_4})$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.121 | 0.000 | 0.739 | 0.970 | |
| $\mathbb{E}(\hat{\pi}_{q_5}+\hat{\pi}_{q_7})$ | 0.000 | 0.000 | 0.000 | 0.003 | 0.168 | 0.185 | 0.739 | 0.832 | |
| $\mathbb{E}(\hat{\pi}_{q_6}+\hat{\pi}_{q_7})$ | 0.000 | 0.000 | 0.000 | 0.001 | 0.100 | 0.044 | 0.543 | 0.659 | |

learning respectively). The approximate 15 percent—based on combined unconditional probabilities for utility functions $V_{g_3}$ and $V_{g_4}$, 0.010 and 0.135 respectively—of respondents who are exogenously identified as showing signs of fatigue are more price sensitive and opposed to all fish species becoming extinct. Compared to learning, the shifts in preferences due to fatigue are more prominent. While this could be interpreted as fatigue having a larger impact than learning, we cannot rule out that it may also be, at least partially, an artifact of a different scale parameter in the two phases.

Although, the interpretations of the other parameters remain largely unchanged from Model 2, it is interesting to point out the differences in the learning and fatigue off-set parameters. In contrast to Model 2, where the fatigue off-set is the only significant off-set parameter, in Model 4 it is the only insignificant off-set. In addition to this, the two other fatigue off-sets are in a different direction compared to Model 2. We remark that this model is associated with a much improved model fit over Model 1. More importantly, we note that this model has a superior fit over Model 2 (an increase by over 300 log-likelihood units at the expense of three additional parameters), where the learning and fatigue patterns are restricted to be the same for everyone. A comparison of the $\bar{\rho}^2$ and information criteria statistics confirm this finding even after accounting for the loss of parsimony. This, along with the differences in learning and fatigue behavior, demonstrate the inappropriateness of the deterministic specification under Model 2.

Compared to Model 3, Model 5 (equation [7]) is a more flexible specification that recognizes the inconsistency in variance may not necessarily be the same for all respondents, but instead can be adequately described by the segmentation described in equation [7a]. Looking firstly at the estimated utility coefficients obtained from this model reveals that they are quite similar to those uncovered in the previous models. But of greater interest is the fact that the unconditional class probability of constant error variance (i.e., $\mu_{s_1}$ in equation [7a]) is 80 percent. This is an important finding, as it highlights that the previous attempts to explain only differences in variance across choice tasks may be erroneous, as they assume that all respondents follow a similar pattern. While the high proportion with the same scale parameter in the three phases is somewhat reassuring, it does imply that over one-fifth of respondents do not comply with the typical assumption of homoscedasticity across the choice sequence. The estimated scale parameter off-sets for the early and late choice tasks (i.e., $\eta_{\mathrm{E}}$ and $\eta_{\mathrm{L}}$ respectively) are relatively large and significant. We remark that these off-sets in variance are in the opposite direction from those estimated in Model 3. We note, however, that the scale off-sets in Model 3 apply to all, respondents, which we demonstrate is probably erroneous. We are also mindful that interpreting the scale parameters in both these models is problematic due to possible confounding with preference heterogeneity. Once more, we find that respondents who are not consistent, are most likely to go through a phase of learning and fatigue. The unconditional probability associated with this class (i.e., $\mu_{s_4}$ in equation [7a]) is 15 percent. Moreover, from Table 2 we find that one-tenth have a mean conditional probability greater than 73 percent for this latent class and in almost one-fifth of cases it is the class that respondents are most likely to belong. While we find no evidence to support the third class (i.e., the 95$^{\text{th}}$ of $\mathbb{E}(\hat{\pi}_{s_3})$ is zero), for more than 7 percent of respondents we find a different scale parameter in the early phase compared the latter two phases of the experiment (over 5 percent have respondent-specific probabilities greater than 50 percent).

We further note that although the model fit for Model 5 is inferior to that attained in Model 4, it is superior compared to Models 1–3. Compared to its deterministic equivalent in Model 3, there is an improvement of almost 150 log-likelihood units at the expense of four additional parameters, which is supported by the improvements in the $\bar{\rho}^2$ and information criteria statistics.

Model 6 (equation [8]) combines elements of Models 4 and 5. Specifically, it accounts for the fact that as respondents proceed there may be a subset whose preferences *or* variance change according to the patterns expressed in equation [8a]. It should also better facilitate interpretation as the off-sets in preferences and error variance are isolated. Under this model we, again, find a high unconditional probability (almost two-thirds) for no signs of learning or fatigue. In this case, however, this includes respondents who have consistent preferences and variance (as depicted by the first utility function, $V_{q_1}$, in equation [8a]). Not surprisingly, the proportion of consistency is smaller than that estimated in the previous PDP models. An approximate equal share (17 percent, based on the unconditional estimates, in each case) of respondents have different preferences or choice variability as they progress through the experiment.

Similar to Model 4, respondents who are identified as having inconsistent preferences are most likely to exhibit both learning and fatigue—combined, based on the means of the conditional probability distributions reported in Table 2, almost one-quarter of respondents are most likely to hold either utility functions $V_{q_4}$ or $V_{q_7}$ (i.e., 0.117 + 0.122). The effects of learning on preferences is, again, found to be more prevalent compared to fatigue, since the 90$^{\text{th}}$ percentile of respondent-specific probabilities for classes $V_{q_2}$ or $V_{q_4}$ (i.e., $\mathbb{E}\left(\hat{\pi}_{q_2} + \hat{\pi}_{q_4}\right)$) is considerably larger than the respective figure for classes $V_{q_3}$ or $V_{q_4}$ (i.e., $\mathbb{E}\left(\hat{\pi}_{q_3} + \hat{\pi}_{q_4}\right)$). Respondents with different preference structures in the early phase of the experiment are found to be significantly more price sensitive and more opposed to the status quo. Respondents identified as having different preferences in the late phase of the experiment are also more price sensitive and opposed to the status quo. A more striking result is the significant reduction in preferences for preserving the fish species as a result of fatigue. In fact, the reduction is to the extent that the actual marginal utilities of this attribute are mostly negative. However, further investigation reveals that the mean of the marginal utility distribution for this subset of respondents is not significantly different from zero. This finding reinforces a previous inference that perhaps there are different processing strategies for attributes during the sequence of choice tasks.

Turning to the subset who have an inconsistent error variance during the experiment, we, again, find that fatigue is more apparent (based on the conditional estimates, 6 percent of respondents are most likely to be described by the fifth utility function in equation [8a]) compared to fatigue, where no respondents are predicted to be associated with (according to the conditional probabilities). The estimated off-sets in the scale parameter indicate that the choice variability for respondents who did not have a constant scale parameter is much lower in early choices compared to those in the middle. We remark that while this provides some supporting evidence for studies that accommodate variance differences for learning and fatigue, it is important to bear in mind that a large proportion of respondent's choices are most likely to exhibit the same variability throughout the experiment. In this regard, the more flexible PDP model presented here seems better equipped to uncover the actual impact of learning and fatigue on error variance, while also isolating its role on preference consistency. This gives it a clear advantage over the earlier models as well as over many of the methods used in earlier studies.

Findings from Model 6 highlight potential confounding between inconsistent preferences and variance and calls for the necessity for specifications that can accommodate both types of inconsistency. We further remark that Model 6 obtains the best model fit. The improvements are substantial. Compared against Models 4 and 5, there is an improvement of 42 and 238 log-likelihood units respectively. Importantly, this improvement in fit is supported by the $\bar{\rho}^2$ and both information criteria statistics even after penalizing for the additional 5 and 9 parameters respectively.

Interestingly, when compared with the behavior patterns summarized in Day et al. (2012,

Table 1, p75), the findings from Model 6 provide an impression of the possible causes of the order effects. Firstly, our main result of changing preference structures for the cost and non-cost attribute as well as for the status quo alternative hint of institutional and/or preference learning. The significant changes also make it difficult to rule out the effects of anchoring and referencing by these respondents. The increase in randomness observed for some respondents as they moved from the early choices to middle choices supports the ideas of an early onset of failing credibility and fatigue effects. Furthermore, the decreased randomness during the late phase would further confirm institutional and/or preference learning occurring at this phase and, importantly, suggest that this type of behavior may even be exhibited by some respondents throughout the entire sequence of choice tasks.

*5.2. Implications for marginal willingness to pay estimates*

From a policy perspective it is important to assess how sensitive the assumptions of preference and variance consistency have on the estimates of marginal WTP for fish conservation. For this, in Figure 1 we present the conditional marginal WTP distributions, as outlined in equation [9b], for all six model specifications. The values represent the expected value of how much respondents would be willing to pay, through a one-off increase in their Income Tax and/or Value Added Tax contributions, for a single fish species to be preserved.

In line with evidence presented elsewhere (e.g., Caussade et al., 2005; Holmes and Boyle, 2005; Campbell, 2007), the histograms suggest that different treatments of learning and/or fatigue has an impact on the derived marginal WTP estimates. Inspecting the distributions, reveals a share with negative marginal WTP in all specifications. This is, obviously, an artifact of our choice of Normal distribution for the fish attribute and the fact that this distribution is fitted to the data (even though the preferences themselves may not actually be Normal). This said, it is interestingly to note the stark difference between Models 1–5 (Figures 1(a) to 1(e))—where the proportion of negative respondent-specific marginal WTP estimates are in the range of 10–20 percent—and our final model (Figure 1(f))—where less than 2 percent of the means of the conditional distributions are negative. The fact that the final model had a superior fit, would suggest that the other models substantially over predict the proportion of the marginal WTP distributions in the negative domain. Bearing in mind that our *a-priori* expectation was that marginal WTP for preserving rare fish species would be positive, this is an important finding.

At the upper tail, we observe that the naïve model that does not address any learning and/or fatigue (i.e., Model 1) and the models that account only for changes in preferences (i.e., Models 2 and 4) produce a high share of extreme respondent-specific marginal WTP estimate. The high marginal WTP values implied by these model cast doubt on their reliability. In fact, in the case of Models 2 and 4, almost 60 percent and 30 percent respectively of these estimates respectively exceed €80, which is perhaps higher than one might expect to pay to conserve a single fish species in a single river catchment.

Figure 1 also shows that the central tendency measures are sensitive to the manner in which learning and fatigue behavior is addressed. Models 1, 3, and 5 produce similar central tendency statistics, all of which are in the region of €25. The equivalent statistic for Model 6 is slightly higher, at approximately €40. However, given that this model gives the best fit and the fewest implausible negative respondent-specific marginal WTP estimates, we can be relatively assured that this model leads to the appropriate marginal WTP distribution.

As a final exploration, in Figure 2 we present back-to-back histograms to compare the means of the conditional marginal WTP distributions against the likelihood of belonging in each latent class in Model 6. The left histogram shows the distribution for respondents whose expected

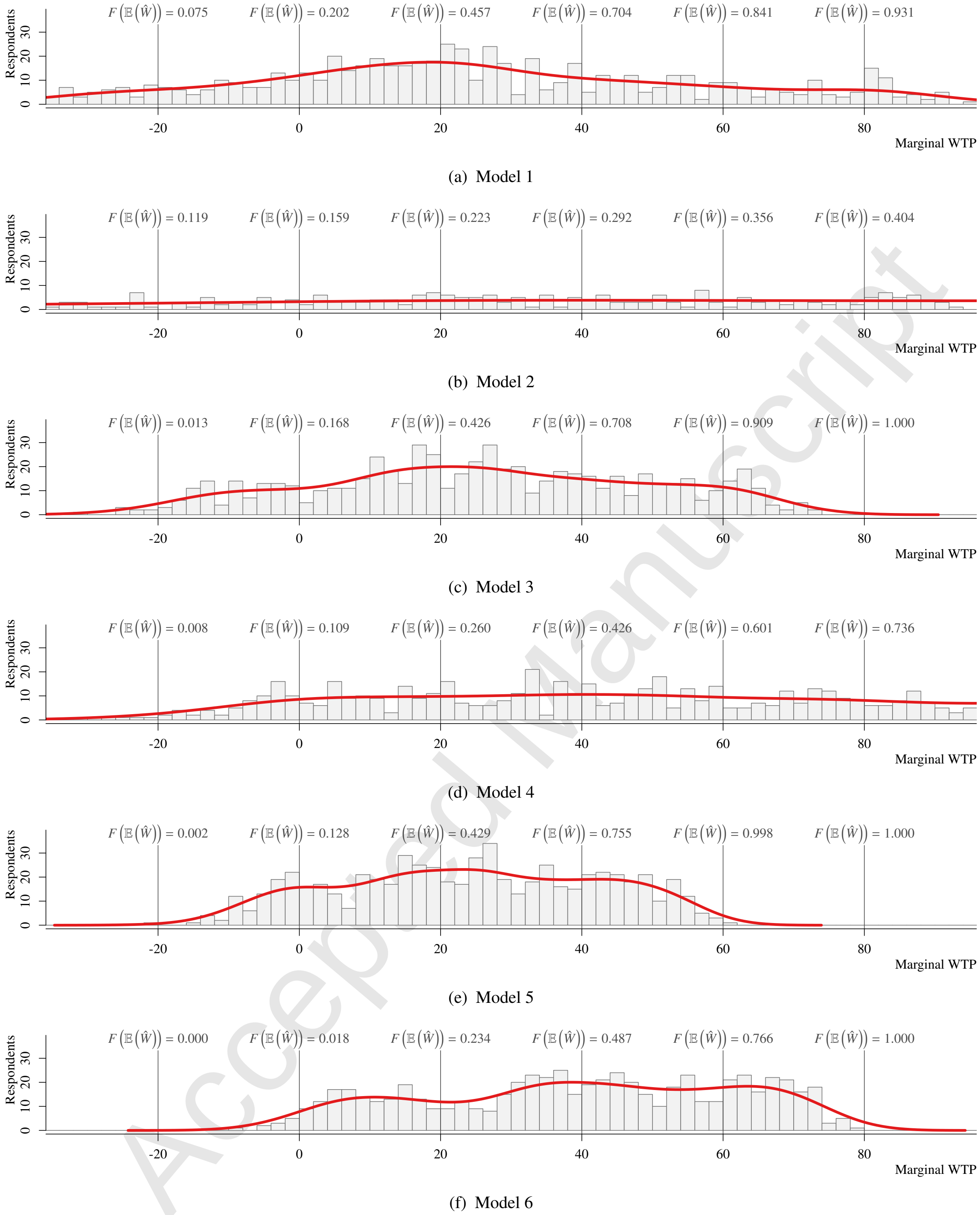


Figure 1: Histograms of the means of the conditional marginal WTP distributions (€)

value is less than or equal to the median of the distribution—i.e., all else being equal, these respondents are the least likely to belong in the class. In contrast, the right histogram is the distribution of the means of the conditional marginal WTP distributions for respondents whose conditional class probability is above the median—*ceteris paribus*, respondents who are most likely to belong in the class.

It is apparent that as one moves from the left to the right histogram for the first latent class

Left histogram depicts respondents with conditional class probability less than or equal to the median conditional class probability (i.e., $\mathbb{E}(\hat{\pi}_n) \leq \widetilde{\mathbb{E}(\hat{\pi})}$).
Right histogram depicts respondents with conditional class probability greater than the median conditional class probability (i.e., $\mathbb{E}(\hat{\pi}_n) > \widetilde{\mathbb{E}(\hat{\pi})}$).

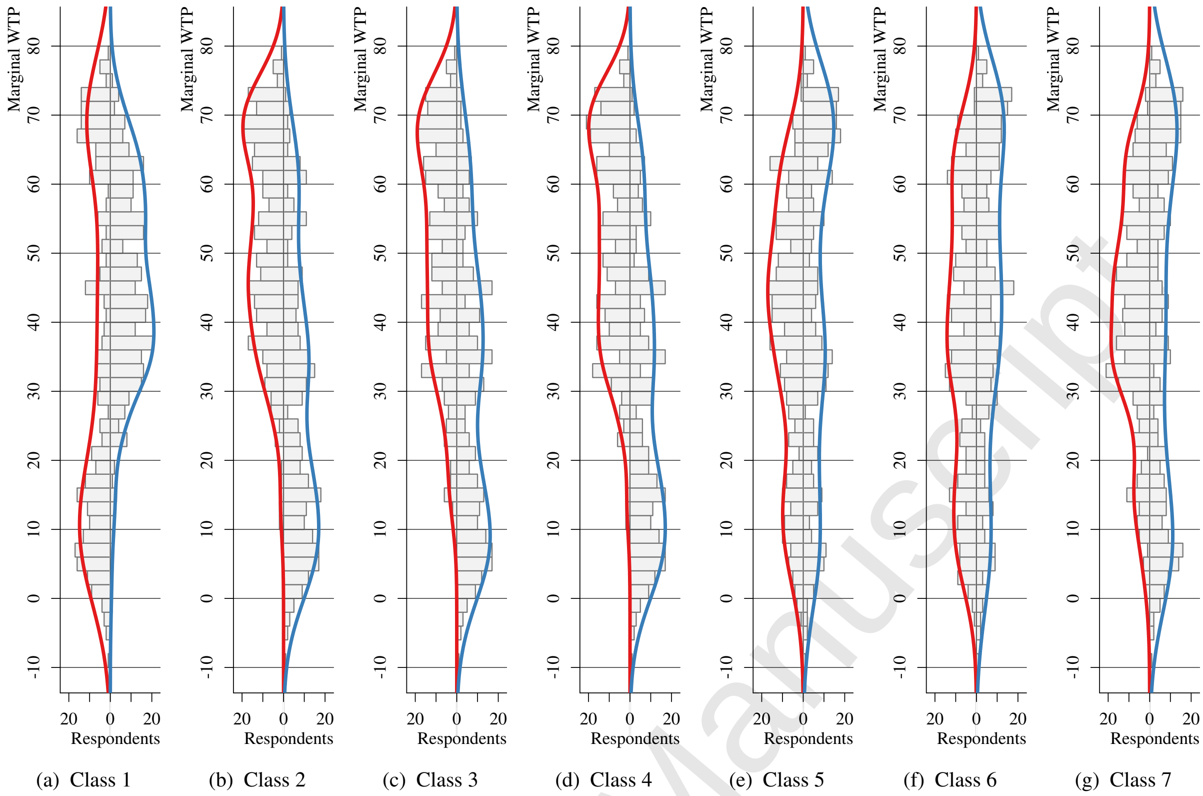


Figure 2: Histograms of the means of the conditional marginal WTP distributions (€) retrieved from Model 6 against conditional class probabilities

(Figure 2(a)) the WTP distribution shifts markedly upwards. This means that respondents who are most likely to exhibit stable preferences and variance (recall that the first class relates to $V_{q_1}$, in equation [8a]) typically are willing to pay more to prevent the fish from becoming extinct. Figures 2(b)–2(d) show that respondents who are predicted as being most likely to hold different preferences in the early and/or late phases have larger probability masses in the lower end of the marginal WTP distribution (e.g., less than €20), which reflects their higher price sensitivity during these choice tasks. It is interesting to note that there is an apparent lower degree of peakedness in the right histograms in Figures 2(e)–2(g). While this could signal a higher degree of preference heterogeneity among respondents most likely to have different choice variability in the learning and/or fatigue phases, care is needed since it is not possible to disentangle these factors.

## 6. Conclusions

In the present paper, we explored the issue of learning and fatigue in the context of inconsistent preferences and variance as respondents progress through the experiment.

Our paper started with the conventional approach of dealing with learning and fatigue, whereby respondents are all treated as either exhibiting inconsistent preferences or inconsistent error variance across early, middle, and late choice tasks. We then moved away from this deterministic method for uncovering learning and fatigue behavior and propose the use of probabilistic decision process (PDP) models (similar in form to latent class models, but where we defined each class to denote a specific learning and fatigue pattern, such as those described in Day et al. (2012, Table 1, p75)). The approach is used to assign respondents into subgroups according to changing preferences and variance parameters at different phases in the experiment to reflect different learning and fatigue behavior. Moreover, we developed a further scale-adjusted PDP model to concurrently accommodate the inconsistent preferences and choice variability at different phases of the experiment. We also estimated these models with random parameters to better capture preference heterogeneity within each phase.

Using a dataset collected to establish marginal willingness to pay (WTP) for the preservation of rare fish species in Ireland we find evidence of both learning and fatigue behavior across the panel of choices. We note that in general we find that the estimated preferences as well as error variances are statistically different between the subset of early, middle and late choice tasks. However, we do note that the incidence of this differs between the six model specifications used in the paper. Unlike all previous studies in both DCEs and behavioral economics that have assumed that the same patterns of inconsistent preferences and variance applies for all respondents, results from our PDP model suggests that this may be an inappropriate assumption. Indeed, our findings suggest that by comparison with the middle, phase of the experiment it may only be a small subgroup of respondents who have different preferences or error variance in the early and late phases of the experiment (approximately 10 percent in both cases).

Comparing the model fit across the six specifications we find that our final model, which aims at accommodating both inconsistent preferences and variance simultaneously, represents the best model fit. Nevertheless, our previous models, which model the impact of learning and fatigue on preferences and error variance separately, suggest somewhat contrary effects related to inconsistent preferences and variance. This highlights potential confounding between the two types of inconsistency and, thus, the necessity for econometric specifications that can accommodate both inconsistent preferences and variance concurrently. Focusing solely on one type of inconsistency may explain only part of the story and, crucially, could lead to biased inferences regarding the impact of learning and fatigue. It is also apparent that this further manifests into clear differences in marginal WTP estimates and choice predictions. Our results reveal that around two-thirds of respondents had consistent preferences and variance throughout the experiment. The remaining respondents are identified as having either inconsistent preferences or choice variability in approximately equal proportions. Incidentally, we find that these respondents resemble the patterns of order effects summarized in Day et al. (2012, Table 1, p75), especially institutional learning, preference learning, failing credibility and fatigue effects. This is important, since it indicates that many different order effects can be analytically determined. Our approach, therefore, offers a way for DCE practitioners to test for such effects, even though they were not envisaged at the survey design stage or during data collection.

In this study, respondents were asked to value quite an unfamiliar, and for the most part a non-use, environmental good and despite this, the majority of respondents (around two-thirds) emerge with consistent preferences and error variance across the different phases of the experiment. It is, nevertheless, of some concern that approximately one-third of respondents did not appear to comply with the usual assumption typically adopted by discrete choice analysts of consistent preferences and error variance. So, although it may only be a small subset of respondents who violate this assumption, our findings reinforce the need to test for learning and fatigue. Overall, our (scale-adjusted) PDP approach provides an intuitive mechanism to explore these phenomena. Our findings provide important information for researchers on how respondents establish their preferences for these types of goods as well as the extent to which respondents understand the valuation exercise at hand and how the cognitive effort may change during the experiment. Armed with this type of information, practitioners designing DCEs should benefit.

Our findings provide, what we feel is compelling, evidence for further investigation into this area. Indeed, we suggest that others replicate this approach on their own data so that we learn even more about the patterns of learning and fatigue in DCEs in a wide variety of settings. Future studies should incorporate procedures for identifying and dealing with respondents who exhibit learning and fatigue behavior so that the sensitivity of model performance as well as marginal WTP estimates to learning and fatigue can be further evaluated. We acknowledge that we focused only on preference and variance consistency and have overlooked any changes in the processing strategies that may have been adopted during the sequence of choices, meaning that we have no way of confirming that our findings are not due to changes in decision rules. An interesting extension to our approach would, therefore, be to use it to explore the incidence of different processing strategies along the sequence of tasks. However, this is likely to increase model complexity and, therefore, pose some estimation and identification problems. Moreover, it will also not completely resolve the confounding issue. We further recognize that although we have not identified the characteristics of respondents who exhibited the various patterns of learning and/or fatigue, we suggest that including socio-economic covariates in the class-membership function is potentially another interesting extension to this modeling approach.

**Appendix A: Alternative definitions of early, middle and late choices**

The results reported in the main paper are derived under the assumption that the first eight choice tasks are early choices and the last eight choice tasks are split equally into middle and late choice tasks. This assumption was reached after comparing the equivalent models and modeling procedures under a range of competing definitions of early, middle and late choice tasks. We appreciate that this is essentially a 'trial and error' process and it may not be feasible in some settings. Unfortunately, however, *a-priori* there is really no way to know the most likely choice tasks where learning ends and fatigue begins. This is why we felt it necessary to conduct and present our analysis of alternative definitions of early, middle and late choices. We present summary results from this in Table A1.

Conditional on there being at least one choice task being categorized into each of the three phases of the experiment for our empirical dataset, where the sequence of choices consisted of 16 choice tasks, there are 105 combinations (i.e., $\sum_{q=1}^{14} q = 105$).[6] In the first instance, we restricted our investigations to the 15 situations where there was a sequence of at least four choice tasks in each phase. While we recognize that this may seem as a somewhat arbitrary decision, we feel that the conditional probabilities of class membership may not have been reasonably reliable if based on panels of less than four choice tasks. Moreover, bearing in mind that we estimated every model with several hundred quasi-random draws and estimated all of the PDP models many times using different starting values, this also had the advantage of restricting the number of models to estimate to a manageable number. We believe that this is sufficient to deal with the matter at hand and, importantly, feel that reducing to a sequence of less than four tasks runs the risk that the results could be unduly influenced by a particular

[6]It is worth mentioning that instead of three phases it would be possible to also facilitate the situation where there are only two phases. In this case the number of combinations would increase to 120 combinations (i.e., $\sum_{q=1}^{15} q = 120$).

Table A1: Summary results of alternative definitions of early, middle and late choices

(a) Early choices: 1–4; Middle choices: 5–8; Late choices: 9–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,361.640 | - | - | - | 1.000 | - | - | - | 25.66 |
| Model 3 | -7,374.292 | - | - | - | - | - | - | 1.000 | 25.38 |
| Model 4 | -7,028.609 | 0.711 | 0.078 | 0.003 | 0.208 | - | - | - | 73.90 |
| Model 5 | -7,357.049 | 0.514 | - | - | - | 0.000 | 0.195 | 0.291 | 24.59 |
| Model 6 | -7,006.497 | 0.602 | 0.068 | 0.018 | 0.167 | 0.061 | 0.022 | 0.061 | 59.19 |

(b) Early choices: 1–4; Middle choices: 5–9; Late choices: 10–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,361.910 | - | - | - | 1.000 | - | - | - | 25.68 |
| Model 3 | -7,369.337 | - | - | - | - | - | - | 1.000 | 25.53 |
| Model 4 | -7,092.342 | 0.733 | 0.073 | 0.000 | 0.194 | - | - | - | 54.71 |
| Model 5 | -7,341.577 | 0.313 | - | - | - | 0.200 | 0.487 | 0.000 | 25.86 |
| Model 6 | -7,080.533 | 0.646 | 0.059 | 0.011 | 0.156 | 0.078 | 0.000 | 0.050 | 44.82 |

(c) Early choices: 1–4; Middle choices: 5–10; Late choices: 11–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,354.991 | - | - | - | 1.000 | - | - | - | 25.90 |
| Model 3 | -7,375.126 | - | - | - | - | - | - | 1.000 | 25.37 |
| Model 4 | -7,294.090 | 0.626 | 0.008 | 0.139 | 0.227 | - | - | - | 13.06 |
| Model 5 | -7,346.027 | 0.000 | - | - | - | 0.210 | 0.790 | 0.000 | 25.75 |
| Model 6 | -7,129.119 | 0.659 | 0.067 | 0.018 | 0.129 | 0.068 | 0.000 | 0.058 | 39.42 |

(d) Early choices: 1–4; Middle choices: 5–11; Late choices: 12–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,359.875 | - | - | - | 1.000 | - | - | - | 26.21 |
| Model 3 | -7,376.902 | - | - | - | - | - | - | 1.000 | 25.39 |
| Model 4 | -7,193.320 | 0.787 | 0.067 | 0.010 | 0.135 | - | - | - | 37.43 |
| Model 5 | -7,347.163 | 0.353 | - | - | - | 0.213 | 0.433 | 0.000 | 24.37 |
| Model 6 | -7,177.533 | 0.672 | 0.055 | 0.022 | 0.091 | 0.085 | 0.000 | 0.075 | 34.38 |

(e) Early choices: 1–4; Middle choices: 5–12; Late choices: 13–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,359.311 | - | - | - | 1.000 | - | - | - | 26.20 |
| Model 3 | -7,373.739 | - | - | - | - | - | - | 1.000 | 25.58 |
| Model 4 | -7,232.999 | 0.817 | 0.067 | 0.021 | 0.096 | - | - | - | 32.93 |
| Model 5 | -7,345.680 | 0.000 | - | - | - | 0.206 | 0.794 | 0.000 | 24.91 |
| Model 6 | -7,215.167 | 0.699 | 0.057 | 0.019 | 0.080 | 0.070 | 0.000 | 0.076 | 30.36 |

(f) Early choices: 1–5; Middle choices: 6–9; Late choices: 10–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,366.722 | - | - | - | 1.000 | - | - | - | 25.48 |
| Model 3 | -7,368.447 | - | - | - | - | - | - | 1.000 | 25.53 |
| Model 4 | -7,053.293 | 0.744 | 0.046 | 0.010 | 0.200 | - | - | - | 62.17 |
| Model 5 | -7,355.634 | 0.448 | - | - | - | 0.000 | 0.252 | 0.300 | 24.67 |
| Model 6 | -7,047.970 | 0.674 | 0.042 | 0.011 | 0.184 | 0.052 | 0.000 | 0.038 | 53.60 |

(g) Early choices: 1–5; Middle choices: 6–10; Late choices: 11–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,359.971 | - | - | - | 1.000 | - | - | - | 25.72 |
| Model 3 | -7,375.098 | - | - | - | - | - | - | 1.000 | 25.38 |
| Model 4 | -7,267.720 | 0.591 | 0.039 | 0.071 | 0.300 | - | - | - | 10.02 |
| Model 5 | -7,369.270 | 0.561 | - | - | - | 0.000 | 0.249 | 0.189 | 25.04 |
| Model 6 | -7,099.343 | 0.696 | 0.050 | 0.016 | 0.147 | 0.043 | 0.000 | 0.047 | 42.90 |

(h) Early choices: 1–5; Middle choices: 6–11; Late choices: 12–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,364.766 | - | - | - | 1.000 | - | - | - | 26.06 |
| Model 3 | -7,377.045 | - | - | - | - | - | - | 1.000 | 25.40 |
| Model 4 | -7,286.167 | 0.669 | 0.059 | 0.065 | 0.207 | - | - | - | 14.60 |
| Model 5 | -7,350.009 | 0.437 | - | - | - | 0.167 | 0.396 | 0.000 | 25.95 |
| Model 6 | -7,149.460 | 0.712 | 0.048 | 0.013 | 0.117 | 0.047 | 0.000 | 0.063 | 35.87 |

(i) Early choices: 1–5; Middle choices: 6–12; Late choices: 13–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,364.269 | - | - | - | 1.000 | - | - | - | 26.03 |
| Model 3 | -7,373.798 | - | - | - | - | - | - | 1.000 | 25.59 |
| Model 4 | -7,305.435 | 0.729 | 0.059 | 0.058 | 0.154 | - | - | - | 17.29 |
| Model 5 | -7,348.140 | 0.000 | - | - | - | 0.165 | 0.835 | 0.000 | 26.01 |
| Model 6 | -7,192.713 | 0.755 | 0.042 | 0.009 | 0.096 | 0.034 | 0.000 | 0.064 | 30.54 |

(j) Early choices: 1–6; Middle choices: 7–10; Late choices: 11–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,358.914 | - | - | - | 1.000 | - | - | - | 25.69 |
| Model 3 | -7,375.006 | - | - | - | - | - | - | 1.000 | 25.37 |
| Model 4 | -7,244.795 | 0.522 | 0.017 | 0.030 | 0.431 | - | - | - | 0.92 |
| Model 5 | -7,335.754 | 0.000 | - | - | - | 0.179 | 0.821 | 0.000 | 24.99 |
| Model 6 | -7,038.929 | 0.685 | 0.047 | 0.018 | 0.124 | 0.046 | 0.000 | 0.080 | 39.81 |

(k) Early choices: 1–6; Middle choices: 7–11; Late choices: 12–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,363.299 | - | - | - | 1.000 | - | - | - | 26.00 |
| Model 3 | -7,376.382 | - | - | - | - | - | - | 1.000 | 25.39 |
| Model 4 | -7,118.008 | 0.791 | 0.045 | 0.012 | 0.151 | - | - | - | 43.30 |
| Model 5 | -7,269.106 | 0.772 | - | - | - | 0.058 | 0.017 | 0.154 | 24.55 |
| Model 6 | -7,095.902 | 0.696 | 0.043 | 0.014 | 0.127 | 0.046 | 0.000 | 0.074 | 38.24 |

(l) Early choices: 1–6; Middle choices: 7–12; Late choices: 13–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,363.188 | - | - | - | 1.000 | - | - | - | 25.97 |
| Model 3 | -7,373.457 | - | - | - | - | - | - | 1.000 | 25.57 |
| Model 4 | -7,300.241 | 0.678 | 0.077 | 0.019 | 0.226 | - | - | - | 12.15 |
| Model 5 | -7,335.807 | 0.000 | - | - | - | 0.178 | 0.822 | 0.000 | 25.06 |
| Model 6 | -7,145.010 | 0.717 | 0.046 | 0.014 | 0.110 | 0.047 | 0.000 | 0.065 | 33.87 |

(m) Early choices: 1–7; Middle choices: 8–11; Late choices: 12–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,362.582 | - | - | - | 1.000 | - | - | - | 25.97 |
| Model 3 | -7,373.747 | - | - | - | - | - | - | 1.000 | 25.33 |
| Model 4 | -7,050.425 | 0.783 | 0.049 | 0.015 | 0.153 | - | - | - | 49.53 |
| Model 5 | -7,247.257 | 0.806 | - | - | - | 0.000 | 0.000 | 0.194 | 24.25 |
| Model 6 | -7,015.997 | 0.669 | 0.049 | 0.015 | 0.126 | 0.063 | 0.000 | 0.079 | 39.89 |

(n) Early choices: 1–7; Middle choices: 8–12; Late choices: 13–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,362.531 | - | - | - | 1.000 | - | - | - | 25.91 |
| Model 3 | -7,371.686 | - | - | - | - | - | - | 1.000 | 25.50 |
| Model 4 | -7,270.171 | 0.652 | 0.082 | 0.008 | 0.258 | - | - | - | 9.20 |
| Model 5 | -7,307.877 | 0.000 | - | - | - | 0.190 | 0.810 | 0.000 | 24.77 |
| Model 6 | -7,071.891 | 0.680 | 0.050 | 0.009 | 0.111 | 0.067 | 0.000 | 0.082 | 35.55 |

(o) Early choices: 1–1; Middle choices: 2–8; Late choices: 9–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,369.765 | - | - | - | 1.000 | - | - | - | 25.29 |
| Model 3 | -7,374.059 | - | - | - | - | - | - | 1.000 | 25.33 |
| Model 4 | -7,145.887 | 0.811 | 0.000 | 0.000 | 0.189 | - | - | - | 39.55 |
| Model 5 | -7,320.395 | 0.815 | - | - | - | 0.030 | 0.009 | 0.147 | 25.43 |
| Model 6 | -7,137.208 | 0.708 | 0.000 | 0.000 | 0.176 | 0.060 | 0.002 | 0.053 | 36.77 |

(p) Early choices: 1–2; Middle choices: 3–8; Late choices: 9–16

| | $\mathcal{L}(\hat{\beta})$ | $\pi_{q_1}$ | $\pi_{q_2}$ | $\pi_{q_3}$ | $\pi_{q_4}$ | $\pi_{q_5}$ | $\pi_{q_6}$ | $\pi_{q_7}$ | WTP |
|---|---|---|---|---|---|---|---|---|---|
| Model 2 | -7,366.609 | - | - | - | 1.000 | - | - | - | 25.21 |
| Model 3 | -7,372.969 | - | - | - | - | - | - | 1.000 | 25.41 |
| Model 4 | -7,121.490 | 0.788 | 0.015 | 0.000 | 0.197 | - | - | - | 43.67 |
| Model 5 | -7,307.296 | 0.840 | - | - | - | 0.000 | 0.027 | 0.133 | 24.23 |
| Model 6 | -7,110.365 | 0.743 | 0.020 | 0.000 | 0.187 | 0.000 | 0.004 | 0.045 | 43.56 |

choice task. To demonstrate that our findings are comparatively consistent and robust under these alternative definitions of early, middle and late phases of the DCE, in Table A1(a)–A1(n) we present summary results from the 14 other definitions. For each definition, we give details of the model fit, class segmentation and marginal WTP predictions obtained from Models 2–6.

A comparison of the log-likelihood values achieved in these definitions of early, middle and late phases reveals that Model 6 provides the best model fit across all definitions. This is followed by Models 4 and 5 respectively and finally Models 2 and 3 respectively. The only exception to this is in the case of Table A1(g), where Model 2 outperforms Model 5.

Given the evidence elsewhere (e.g., Carlsson et al., 2012; Hess et al., 2012; Czajkowski et al., 2014) that the first one or two choices are perhaps the most likely to produce different value estimates and scale parameters, in Tables A1(o) and A1(p) we consider two specifications where the early phase is defined as the first and first two choice tasks respectively. Once more, a general improvement in model fit is evident as one moves from Model 2 to Model 6 in both these settings.

Findings in Table A1 provide further confirmation that recognizing the inconsistent preferences as respondents progress through the choice tasks has the largest impact on model fit. Nevertheless, across all these definitions of early, middle and late choices (including those that defined the first one and two choices as early), the results suggest that addressing the changes in both respondent's preferences and variance at different phases of the experiment is warranted. In general, we find that definitions associated with a long learning phase (up to 8 choice tasks) and a short fatigue phase (as few as the final four choice tasks) produce superior model fits. Nevertheless, we are (understandably) reluctant to recommend this as a 'rule of thumb', since definitions of learning and fatigue need to be considered on a case-by-case basis. What we can say though is that, in the first instance, practitioners may want to consider increasing the number of sequential choice tasks in each phase and/or increase the number of distinct phases, since this will reduce the number of scenarios to test. Similarly, model searches using less elaborate specifications (e.g., with fewer PDP latent classes, fewer random draws, assuming preference homogeneity) can be estimated relatively quickly. This is likely to give analysts a good idea of the appropriate classification of learning and fatigue choice tasks, from which they can focus their efforts.

Irrespective of the classification of early, middle and late choices, the estimated proportion (based on the unconditional estimates) of respondents who are identified as having consistent preferences and consistent variance is generally in the region of 60–70 percent (which is in accordance with those reported in Table 1). While the marginal WTP values reported in Table A1 are estimated using the means of the estimates random parameters, they do, nevertheless, permit comparisons to be made. We note that the magnitudes of marginal WTP estimates do resemble those reported in the main paper. This is an important, and reassuring, finding, as it indicates that the conclusions reached in the main paper generally apply irrespective of the definition used to classify early, middle and late choices.

## References


Alberini, A., 2012. Repeated questioning in choice experiments: are we improving statistical efficiency or getting respondents confused? Journal of Environmental Economics and Policy 1 (2), 216–233.

Ariely, D., Loewenstein, G., Prelec, D., 2003. "Coherent arbitrariness": stable demand curves without stable preferences. Quarterly Journal of Economics 118 (1), 73–106.

Bateman, I. J., Burgess, D., Hutchinson, W. G., Matthews, D. I., 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. Journal of Environmental Economics and Management 55 (2), 127–141.

Bradley, M., Daly, A., 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21, 167–184.

Braga, J., Starmer, C., 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. Environmental and Resource Economics 32, 55–89.

Brouwer, R., 2012. Constructed preference stability: a test-retest. Journal of Environmental Economics and Policy 1 (1), 70–84.
Brouwer, R., Dekker, T., Rolfe, J., Windle, J., 2010. Choice certainty and consistency in repeated choice experiments. Environmental and Resource Economics 46, 93–109.
Campbell, D., 2007. Willingness to pay for rural landscape improvements: combining mixed logit and random-effects models. Journal of Agricultural Economics 58 (3), 467–483.
Campbell, D., Erdem, S., 2015. Position bias in best-worst scaling surveys: a case study on trust in institutions. American Journal of Agricultural Economics 97 (2), 526–545.
Campbell, D., Hensher, D. A., Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. Journal of Environmental Planning and Management 54 (8), 1061–1076.
Campbell, D., Hensher, D. A., Scarpa, R., 2012. Cost thresholds, cut-offs and sensitivities in stated choice analysis: identification and implications. Resource and Energy Economics 34 (3), 396–411.
Campbell, D., Hensher, D. A., Scarpa, R., 2014. Bounding WTP distributions to reflect the 'actual' consideration set. Journal of Choice Modelling 11, 4–15.
Campbell, D., Hess, S., Scarpa, R., Rose, J. M., 2010. Accommodating coefficient outliers in discrete choice modelling: a comparison of discrete and continuous mixing approaches. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-art and the State-of-practice — Proceedings from the Inaugural International Choice Modelling Conference. Emerald Group Publishing, Bingley, pp. 331–352.
Carlsson, F., Martinsson, P., 2001. Do hypothetical and actual marginal willingness to pay differ in choice experiments? Journal of Environmental Economics and Management 41, 179–192.
Carlsson, F., Mørkbak, M. R., Olsen, S. B., 2012. The first time is the hardest: a test of ordering effects in choice experiments. Journal of Choice Modelling 5 (2), 19–37.
Caussade, S., Ortúzar, J. D., Rizzi, L. I., Hensher, D. A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B 39 (7), 621–640.
Czajkowski, M., M. Giergiczny, M., Greene, W. H., 2014. Learning and fatigue effects revisited: investigating the effects of accounting for unobservable preference and scale heterogeneity. Land Economics 90 (2), 324–351.
Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R., Wang, P., 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63 (1), 73–91.
Day, B., Prades, J. P., 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management 59 (3), 271–285.
DeShazo, J. R., Fermo, G., 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 2002, 123–143.
Doornik, J. A., 2007. Object-oriented matrix programming using Ox. Timberlake Consultants Press and Oxford, London.
Erdem, S., Campbell, D., Hole, A. R., 2015. Accounting for attribute-level non-attendance in a health choice experiment: does it matter? Health Economics, in press.
Erdem, S., Campbell, D., Thompson, C., 2014. Elimination and selection by aspects in health choice experiments: prioritising health service innovations. Journal of Health Economics 38, 10–22.
Fosgerau, M., Hess, S., 2009. A comparison of methods for representing random taste heterogeneity in discrete choice models. European Transport 42, 1–25.
Greene, W. H., Hensher, D. A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37, 681–698.
Greene, W. H., Hensher, D. A., 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Applied Economics 45 (14), 1897–1902.
Hensher, D. A., Collins, A. T., Greene, W. H., 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. Transportation 40 (5), 1003–1020.
Hensher, D. A., Greene, W. H., 2003. The mixed logit model: the state of practice. Transportation 30, 133–176.
Hensher, D. A., Rose, J. M., Greene, W. H., 2005. Applied choice analysis: a primer. Cambridge University Press, Cambridge.
Hess, S., 2010. Conditional parameter estimates from mixed logit models: distributional assumptions and a free software tool. Journal of Choice Modelling 3 (2), 134–152.
Hess, S., Bierlaire, M., Polak, J. W., 2005. Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A 39, 221–236.
Hess, S., Hensher, D. A., Daly, A., 2012. Not bored yet–revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: Policy and Practice 46 (3), 626–644.

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

Holmes, T. P., Boyle, K. J., 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics 81, 114–126.

Hu, W., Adamowicz, W. L., Veeman, M. M., 2006. Labeling context and reference point effects in models of food attribute demand. American Journal of Agricultural Economics 88 (4), 1034–1049.

Kingsley, D. C., Brown, T. C., 2010. Preference uncertainty, preference learning, and paired comparison experiments. Land Economics 86 (3), 530–544.

Lagerkvist, C. J., Carlsson, F., Viske, D., 2006. Swedish consumer preferences for animal welfare and biotech: a choice experiment. AgBioForum: The Journal of Agrobiotechnology Management and Economics 9, 51–58.

Lancaster, K. J., 1966. A new approach to consumer theory. The Journal of Political Economy 74 (2), 132–157.

Louviere, J. J., Hensher, D. A., Swait, J. D., 2003. Stated choice methods: analysis and applications. Cambridge University Press, Cambridge.

Magidson, J., Vermunt, J. K., 2008. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. Proceedings Of The Sawtooth Software Conference.

Manski, C. F., 1977. The structure of random utility models. Theory and Decision 8, 229–254.

McNair, B. J., Hensher, D. A., Bennett, J., 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model. Environmental and Resource Economics 51 (4), 599–616.

Oppewal, H., Morrison, M., Wang, P., Waller, D., 2010. Preference stability: modeling how consumer preferences shift after receiving new product information. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-art and the State-of-practice — Proceedings from the Inaugural International Choice Modelling Conference. Emerald Group Publishing, Bingley, pp. 499–516.

Plott, C. R., 1996. Rational individual bebaviour in markets and social choice processes: the discovered preference hypothesis. In: Arrow, K. J., Colombatto, E., Perlman, M., Schmidt, C. (Eds.), The Rational Foundations of Economic Behaviour. Macmillan Press, London.

Plott, C. R., Zeiler, K., 2005. The willingness to pay–willingness to accept gap, the "endowment effect", subject misconceptions, and experimental procedures for eliciting valuations. American Economic Review 95 (3), 530–545.

Sælensminde, K., 2001. Inconsistent choices in stated choice data: use of the logit scaling approach to handle resulting variance increases. Transportation 28, 269–296.

Savage, S. J., Waldman, D., 2008. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. Journal of Applied Econometrics 23, 351–371.

Scarpa, R., Campbell, D., Hutchinson, W. G., 2007. Benefits estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment. Land Economics 83, 617–634.

Scarpa, R., Gilbride, T. J., Campbell, D., Hensher, D. A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. European Review of Agricultural Economics 36 (2), 151–174.

Swait, J., Adamowicz, W., 2001. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational Behavior and Human Decision Processes 86, 141–167.

Thurstone, L. L., 1927. A law of comparative judgment. Psychological Review 34, 273–286.

Yu, J., Vishwanathan, S. V. N., Günter, S., Schraudolph, N. N., 2010. A quasi-newton approach to nonsmooth convex optimization problems in machine learning. Journal of Machine Learning Research 11, 1145–1200.