Article:

Hess, S and Giergiczny, M (2015) Intra-respondent heterogeneity in a stated choice survey on wetland conservation in Belarus: First steps towards creating a link with uncertainty in contingent valuation. Environmental and Resource Economics, 60 (3). 327 - 347. ISSN 0924-6460

https://doi.org/10.1007/s10640-014-9769-9

Intra-respondent heterogeneity in a stated choice survey on wetland conservation in Belarus: first steps towards creating a link with uncertainty in contingent valuation

Stephane Hess,
Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk,
Tel: +44 (0)113 34 36611

Marek Giergiczny,
University of Warsaw, Faculty of Economic Sciences, mgiergiczny@wne.uw.edu.pl,
Tel: +48 22 55 49 157

Abstract

Applications of discrete choice models in environmental valuation increasingly use a random coefficient specification, such as mixed logit, to represent taste heterogeneity. The majority of applications rely on data containing multiple observations for each respondent, where a common assumption is that tastes stay constant across choices for the same respondent. We question this assumption and make use of a model developed in the transport field which allows tastes to vary over choices for each consumer in addition to variation across consumers. An empirical analysis making use of a stated choice dataset for wetland conservation in Belarus shows that superior performance is obtained by allowing jointly for the two types of heterogeneity and that recovery of these intra-respondent variations is not possible using standard approaches, such as allowing for scale heterogeneity across tasks. We show also that intra-respondent heterogeneity can be especially high for attributes which respondents are unfamiliar with, and that a failure to account for it can substantially

affect welfare estimates. We interpret this as an indication that this heterogeneity relates primarily to uncertainty. Finally, we offer initial insights into the relationship between intra-respondent heterogeneity and findings on uncertainty in a contingent valuation context.

Keywords: stated preference data*;* random taste heterogeneity; mixed logit; intra-respondent heterogeneity; wetland conservation

1. Introduction

A key interest in the discrete choice modelling literature is the representation of heterogeneity in sensitivities across respondents, especially in the form of unexplained (or random) differences, typically using the random coefficients (as opposed to error components) specification of the Mixed Multinomial Logit (MMNL) model[1]. The fact that the widely used stated choice (SC) surveys typically collect multiple observations for each respondent can substantially aid the recovery of such random heterogeneity, and the majority of MMNL applications do indeed rely on SC data. The typical assumption in such cases is that tastes/sensitivities vary across respondents, but stay constant across replications for the same respondent (cf. Revelt & Train, 1998).

While empirical evidence clearly supports the notion that accommodating heterogeneity across respondents is more reasonable than assuming that heterogeneity is across all

[1] See for example: McFadden & Train (2000), Hensher & Greene (2003), Train (2003). Also see Hoyos (2010) for an overview of MMNL applications into environmental valuation. While the random coefficients specification focusses on random variations in marginal utility parameters, the error components specification primarily aims to capture correlation across alternatives or heteroscedasticity. The two specifications are mathematically equivalent.

choices, it should be noted that the two are not mutually exclusive. Indeed, a number of papers in transport research have put forward the use of a MMNL structure which allows for random heterogeneity both across individual respondents as well as across choices for the same respondent. In the work by Bhat and Castelar (2002), Bhat and Sardesai (2006), Hess and Rose (2009), Cherchi et al. (2009), Yañez et al. (2011) and Hess & Train (2011), empirical evidence suggests that the incorporation of intra respondent heterogeneity can lead to further gains in fit while however still suggesting that the majority of heterogeneity relates to inter-respondent heterogeneity.

To the best of our knowledge, this combined inter- and intra-respondent specification of the MMNL model has not been applied outside of transport thus far. The value of environmental goods/services in stated preference studies is derived through creating hypothetical markets that people are often unfamiliar with. In such situations, it is difficult to expect that respondents have a-priori well-formed preferences, and this heightens the interest in the study of intra-respondent heterogeneity for the present paper. More importantly however, while evidence of intra-respondent heterogeneity is *interesting* per se, the above body of work has done little to investigate the specific *causes* or *nature* of this within respondent variation in sensitivities. This means that the findings from such research can be of limited benefit especially to practitioners.

Developing a suitable tactic to accommodating intra-respondent heterogeneity requires us first to understand the potential drivers of such variations. Bateman et al. (2008) identify three different concepts of individuals' preferences:

(i) *a-priori* well-formed or readily divined through a single incentive compatible question;

(ii) learned or 'discovered' through a process of repetition and experience; or

(iii) internally coherent but liable to be strongly influenced by some initial arbitrary anchor.

There is a large body of work indicating that (i) is a rather unrealistic assumption, with evidence from the contingent valuation (CV) literature, e.g. Cameron and Quiggin (1994), McFadden (1994) and Bateman et al., (2001), consistently reporting differences in WTP estimated in the first and the second stages in the double bound dichotomous choice format. This is consistent also with findings from empirical examples using MMNL models allowing jointly for inter-respondent and intra-respondent heterogeneity on SC data (see summary in Hess & Train, 2011). The key issue remains the interpretation of such findings.

Discovered preference hypothesis (DPH) (ii) assumes that agents have stable and context free preferences that exist independently of the discovery process (Plott 1996). The role of learning in this context within a SC survey is that individuals answering a sequence of choices discover the best way to act about their preferences. An extensive body of empirical work has investigated how respondents learn to cope with the choice tasks (institutional learning) or discover their preferences through practice and repetition (value learning), with examples in Holmes and Boyle, (2005), Hanley et al. (2005), Brouwer et al. (2010). Other work has focussed more on changes in response quality (or error variance) over the course of a survey, linking improvements to learning and reductions to boredom/fatigue (see e.g. Bradley & Daly, 1994, Adamowicz et al. 1998, and Hess et al., 2012).

An alternative to DPH is that preferences are not discovered during the choice sequence but are constructed (e.g. Slovic 1995; Ariely et al. 2003, 2006). For example Ariely et al. (2003) in their theory of Coherent Arbitrariness (CA) (iii) argue that preferences are path dependent; an arbitrary selected initial value, with an individual's desire to maintain internal consistency, will preserve this anchoring effect through the sequence of choices. CA predicts that respondents will have stable WTP values, or rather a range of acceptable values, which will however be heavily influenced by the initial anchor. In recent years, a growing number of CV applications have provided empirical support to the idea that many people are uncertain about their exact WTP, and may have a WTP range rather than a point value (Hakansson 2008, Ellingson et al. 2009, Hanley et al. 2009, Mahiue et al. 2012). In this context, it is worth noting that answers from a single open-ended CV question are a static representation of respondent uncertainty, which differs somewhat from the dynamic effects in CA, where choices are consistent over time, leading to reduced randomness with more decisions.

Independent of the theory at hand, a distinction arises between a respondent's actual WTP and the reported WTP; indeed, it is conceivable that a respondent does have a point value WTP but chooses to only reveal it as a range. To this extent, the uncertainty or variability is at the analyst end rather than the respondent end. On the other hand, it is similarly possible that the *actual* WTP varies as a function of the specific choice scenario faced, and this substantially increases the scope for such variations in both SC and CV data. Finally, simple respondent uncertainty is also likely to play a role.

It should be acknowledged that the topic of preference uncertainty has been extensively studied in CV even prior to work using ranges of WTP, mainly in the context of the dichotomous contingent valuation format (DC-CV), focussing on the use of certainty scales (e.g. Li and Mattson 1995; Welsh and Poe 1998; see Kobayashi et al., 2012 for a detailed overview of this approach), an approach since also adopted in many SC studies (e.g. Norwood, 2005; Hensher et al., 2011; Lundhede et al., 2009; Dekker et al. 2013 - see also Dekker, 2012 for detailed overview of preference uncertainty in SC). The nature of SC studies (i.e. multi-alternative and multi-attribute trade-offs spread across several scenarios) however creates difficulties in directly interpreting the expressed decision certainty in terms of WTP-certainty.

The phenomena described above have a number of quite distinct potential impacts on results. Learning and fatigue could over the course of a set of SC scenarios lead to changes in both relative sensitivities and absolute sensitivities (i.e. scale). In both cases, one would expect trends to manifest themselves, rather than random fluctuations, and these in turn can be accommodated in a deterministic manner. The key interest in accommodating random intra-respondent heterogeneity lies in capturing less structured variation, i.e. that without a clear trend or dynamic nature. The present paper makes the case that evidence of intra-respondent heterogeneity in models estimated on SC data may well reflect the presence of respondent uncertainty[2]. This is especially true in cases where the intra-

[2] Our empirical work shows that the retrieved intra-respondent heterogeneity is not systematic across tasks and does not seem to relate to either learning or fatigue. This makes the case for the specific random treatment we use, and the interpretation of the variation as linking to uncertainty.

respondent heterogeneity is in alternative specific constants which thus captures heteroscedasticity (analogous to an error components specification) as is the case in our empirical application.

The contribution of the present paper is twofold. First, it is to the best of our knowledge the first application of a MMNL model that allows jointly for inter-respondent and intra-respondent heterogeneity in the context of a SC study on environmental goods valuation[3]. Secondly, it makes a first step in creating a link between findings from SC and CV, using data from a parallel CV component. The assumption we make is that WTP ranges observed in CV at least in part reflect respondents' inability to determine one true point of value/utility/willingness-to-pay. We assume that this uncertainty should also be present in SC. By making a link between the uncertainty ranges reported in the CV part and intra-respondent heterogeneity in SC, our study adds a specific reason why homogenous sensitivities across choices for the same respondent in MMNL models are likely to be too restrictive. It also allows us to move away from simply stating that intra-respondent heterogeneity exists and enables us to link it to a possible cause in the form of respondent uncertainty. We stress that we only offer a first step in making such links between CV and SC, and do not necessarily relate it to a specific existing theory. It is also important to acknowledge that our CV component uses a simple open-ended approach, and there is significant scope for follow up work to our study. An example of more in-depth discussions and more refined CV approaches is the work of Kobayashi et al. (2012).

[3] Going much further than for example the Baerenklau and Barenklau (2005) work which allows for serial correlation in the random coefficients through an AR1 process.

The remainder of this paper is organised as follows. Section 2 presents a brief overview of mixed logit methodology, with a focus on the joint treatment of inter-respondent and intra-respondent heterogeneity. Section 3 presents the data used for our analysis. This is followed in Section 4 by the empirical work, making use of the range of techniques alluded to above. Finally, Section 5 presents the conclusions of the paper.

2 Mixed logit methodology

The MMNL model accommodates taste heterogeneity in a continuous specification through integration of MNL choice probabilities over the assumed multivariate distribution of the vector of taste coefficients β. In particular, let $P_{n,t}(\beta)$ be the MNL probability of the observed choice for respondent *n* in choice situation *t,* conditional on a vector of taste coefficients $\beta.$ The log-likelihood (LL) function of the corresponding MMNL model would then be given by:

$$LL(\Omega) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} ln \int_{\beta} P_{n,t}(\beta) f(\beta|\Omega) \mathrm{d}\beta, \qquad (1)$$

where *N* is the number of respondents, T_n is the number of choice tasks faced by respondent *n,* and the vector of taste coefficients β follows a random distribution $f(\beta|\Omega)$ with a vector of parameters Ω. The log-likelihood function in Equation (1) has no closed form solution, and the typical approach to estimation is to replace the log-likelihood by the simulated log-likelihood (SLL), with:

$$SLL = \sum_{n=1}^{N} \sum_{t=1}^{T_n} ln \sum_{r=1}^{R} \frac{P_{n,t}(\beta_{r,t,n})}{R}, \qquad (2)$$

where $\beta_{r,t,n}$ now represents one of R draws for respondent n in choice situation t from the distribution $f(\beta|\Omega)$. Increasing the number of draws from the multivariate vector β and especially their coverage of the multi-variate domain of β, reduces the error introduced in estimation by using Equation (2) as an approximation of Equation (1).

This specification of the MMNL model is directly applicable to cross-sectional data, where $T_n=1$ for all respondents. For the estimation of MMNL models on repeated choice data, the approach put forward by Revelt & Train (1998) has now become the standard specification. This replaces Equations (1) and (2) by:

$$LL(\Omega) = \sum_{n=1}^{N} ln\left(\int_{\beta} \prod_{t=1}^{T_n}\left[P_{n,t}(\beta)\right] f(\beta|\Omega)\mathrm{d}\beta\right), \tag{3}$$

and

$$SLL = \sum_{n=1}^{N} ln\left(\sum_{r=1}^{R} \frac{\prod_{t=1}^{T_n}\left[P_{n,t}(\beta_{r,n})\right]}{R}\right) \tag{4}$$

respectively, where just a single set of R draws is now used for each respondent n.

This specification is based on an assumption of intra-respondent homogeneity; i.e. sensitivities vary across respondents but stay constant across choices for the same respondent.

A more flexible specification uses integration in two places. It specifies the vector β as a sum of two components, with the sensitivity for person n in task t given by $\beta_{n,t} = \alpha_n + \gamma_{n,t}$, where α varies across respondents but stays constant across choices for the same respondent, while γ varies across all choices. For identification, the mean sensitivities are captured in α, such that the means of the individual elements in the vector γ are all constrained to zero. We have that $\alpha \sim g(\alpha|\Omega_\alpha)$ and $\gamma \sim h(\gamma|\Omega_\gamma)$, and the log-likelihood function for this model is given by:

$$LL(\Omega) = \sum_{n=1}^{N} ln\left(\int_\alpha \left(\prod_{t=1}^{T_n}\left[\int_\gamma P_{n,t}(\alpha,\gamma)h(\gamma|\Omega_\gamma)\mathrm{d}\gamma\right]\right) g(\alpha|\Omega_\alpha)\mathrm{d}\alpha\right), \quad (5)$$

A model of a form such as in Equation (5) has been used in various specifications by Bhat and Castelar (2002), Bhat and Sardesai (2006) and Hess and Rose (2009) and differs somewhat from the specification in Cherchi et al. (2009) and Yañez et al. (2011).

A key issue with models allowing jointly for inter-respondent and intra-respondent heterogeneity is the estimation and in particular the specification of the simulated log-likelihood, which, as discussed in detail in Hess & Train (2011), uses averaging at two different levels, with:

$$SLL = \sum_{n=1}^{N} ln\left(\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_n}\frac{1}{K}\sum_{k=1}^{K} P_{n,t}\left(\alpha_r, \gamma_{k,t}\right)\right)\right). \quad (6)$$

This simulation uses *R* draws of α for respondent *n*, along with KT_n draws of γ, where each draw of γ is used with each draw of α, leading to a total number of RKT_n evaluations of logit probabilities for respondent *n*, compared to just RT_n in the cross-sectional and panel specifications.

3. Survey work

3.1. The policy site

The Polesia region of Belarus was once a land of vast pristine mires and bogs. Nowadays it has only a few large wetlands – a result of ambiguous draining programmes implemented in Soviet times. The remaining wetlands are, however, still quite extensive and relatively intact by European standards. Nevertheless, an on-going natural succession is transforming the wetlands, covering them with bushes and trees and thus eliminating the open and undisturbed space that provides a unique habitat for a number of rare bird species. The Zvanets mire covers almost 16,500 ha and is located in South-Western Belarus, close to the Ukrainian border. Of this, about 10,500 ha are currently protected as a state biological reserve.

For centuries, the mire was used by local farmers for harvesting biomass to feed cattle. As a result of regular and extensive land use, a unique ecosystem emerged. It became a site of international importance and a valuable habitat for a number of rare flora and fauna species. One third of the world population of globally endangered Aquatic Warbler *(Acrocephalus Paludicola L.)* nests here making this area the world's largest breeding site for the species, and crucial to its survival.

A protection management programme could prevent or mitigate the undesirable succession taking place in the Zvanets. The program of annual biomass harvests of 1,500-2,000 ha of the fen mire – with plots alternating every year, so that each place is harvested every few years – is expected to effectively slow down the expansion of shrubs. Four management scenarios have been proposed: hand scythe mowing, *mechanical* mowing, controlled burning of the dry biomass in winter and *chemical* treatment of shrubs with herbicides[4].

3.2. Questionnaire structure and sampling of respondents

---

[4] For more detailed description of the programmes and the policy site see Valasiuk et al. (2013).

To understand the value of preserving the wetlands, a survey was conducted in January 2010. The questionnaire was administered face-to-face on a broad sample of the Belarusian population, with interviews conducted in respondents' houses. Questionnaires were randomly assigned to 270 individuals and 206 valid questionnaires were collected.

The questionnaire consisted of five parts. The first one included questions about respondents' attitudes towards biodiversity and conservation issues. The second part described the ecological importance of stopping the succession of trees and bushes and introduced possible policy options. Since the protection of the Zvanets mire is important for saving the flagship Aquatic Warbler *(Acrocephalus Paludicola L.)*, maps with its current distribution, breeding sites and photos were presented. The fourth part of the survey contained the CV and SC tasks. Finally, the fifth part contained debriefing questions and collected socio-economic data, including gender, age, location, education, household characteristics and income.

3.3. CV component

Each respondent was asked the classic and interval open-ended (CIOE) question proposed by Håkansson (2008), providing either an exact amount or an interval valuation. All respondents were asked to report their WTP for the same active protection programme, using mechanical mowing of 2,000 ha/year and enlarging the Zvaniets reserve by 4,000 ha. The following wording of the CIOE question was used:

Try to state what you are willing to pay for the described protection programme, either as an interval between two amounts or as an exact amount.

Answer: (fill in ONE of the options below) **Option 1:** I am willing to pay between .... and ………. this year as a lump sum. **Option 2:** I am willing to pay .... this year as a lump sum.

The order of the two options was rotated across respondents. The share of respondents choosing to report a range rather than a single point value is lower than in Hanley et al. (2009), but still remains high, at 66 per cent.

3.4. Stated choice component

Each respondent was next faced with 16 choice situations on a computer screen, involving the choice between the status quo (SQ) alternative, with no protection programme and no payment required, and three programme alternatives. These were described in terms of the method of removing shrubs, the area where shrubs are removed, the enlargement of the reserve, and the cost of the protection programme. The payment vehicle used in the survey was an obligatory annual payment that all Belarusian residents would have to make to a fund exclusively dedicated to the protection of the Zvanets wetland.

Table 1 presents the attributes and their levels used in the questionnaire, which were determined through a process of consultation with policy-makers and biologists. The choice sets were prepared following the Street et al. (2005;2007) optimal-in-difference design approach. The tasks were assigned to respondents in a random order, meaning that any variations in sensitivities we see across tasks are not the result of the experimental design. Despite capturing data using a best-worst-best approach, we used only the data on best choice, in part given recent findings by Giergiczny et al. (2013) and Rose (2014) relating to inconsistencies between the preferences leading to best and worst choices.

Table 1. Attributes and attribute levels used in the stated choice scenarios

Attributes	Levels
Method of removing shrubs[5]	None (SQ) Hand scythe mowing (*manual*) Mechanical mowing (*mechanical*) Controlled burning of the dry biomass in winter (*burning*) Using herbicides (*chem*)
Area where shrubs are removed (*area*)	0 ha / year (SQ) 1 000 ha / year 2 000 ha / year 3 000 ha / year 4 000 ha / year
Enlargement of the reserve (*reserve*)	0 ha (SQ) 2 000 ha 4 000 ha 6 000 ha
Cost (*cost*)	0 BYR[6] (SQ) 5 000 BYR 30 000 BYR 55 000 BYR 80 000 BYR

[5] The variables were effects-coded. Respondents were informed about the pros and cons of each technique.

[6] BYR – Belarusian Ruble, The exchange rate in Jan 2010, 1 Euro = 4 200 BYR.

4. Empirical analysis

4.1. Base model analysis

4.1.1. Specification

Following initial investigations which did not reveal consistent and significant nonlinearities in response with the data at hand, a purely linear in attributes specification was used. These findings also mean that we can reject non-linearities as a possible reason for the intra-respondent heterogeneity found in the MMNL models. For *area* and *reserve*, we worked in thousands of ha in the utility function, while for *cost*, we worked in thousands of BYR.

Four different MNL models[7] were estimated in the first stage of our empirical work. The first (1) is a simple structure with generic parameters across the sixteen tasks and assumes scale homogeneity across tasks, i.e. keeping the variance of the extreme value term constant across replications. The model uses effects coding for the method of shrub removal, a constant for the SQ option, and linear effects for the size of the *area* where shrubs are removed, the amount of enlargement of the *reserve* and the *cost*. This is followed by three

[7] We used MNL models here as the cross-sectional focus on structural intra respondent heterogeneity would conflict with the panel nature of the MMNL model.

departures from this model. We first (2) allow for scale differences across the sixteen tasks (using the first task as the base) before moving to task specific models (3) , i.e. allowing also for variations in relative sensitivities. Finally, we test a constrained version of this model, focussing on variation across stages separately for each of the seven estimated parameters, i.e. using seven separate models (4-10).

4.1.2. Testing for structural changes in preferences

The results from this process are summarised first in Table 2 for model fit statistics. We note that the model with separate scale parameters (2) as well as the model with task specific estimates for all parameters (3) fail to offer statistically significant improvements over the base model, using a simple likelihood ratio (LR) test. Turning to the models (4-10) which look at individual parameters in turn[8], we note that only for *manual* do we see significant improvements in fit at usual levels of confidence when using task specific estimates, where the improvements when doing the same for *cost* are only significant at the 90% level.

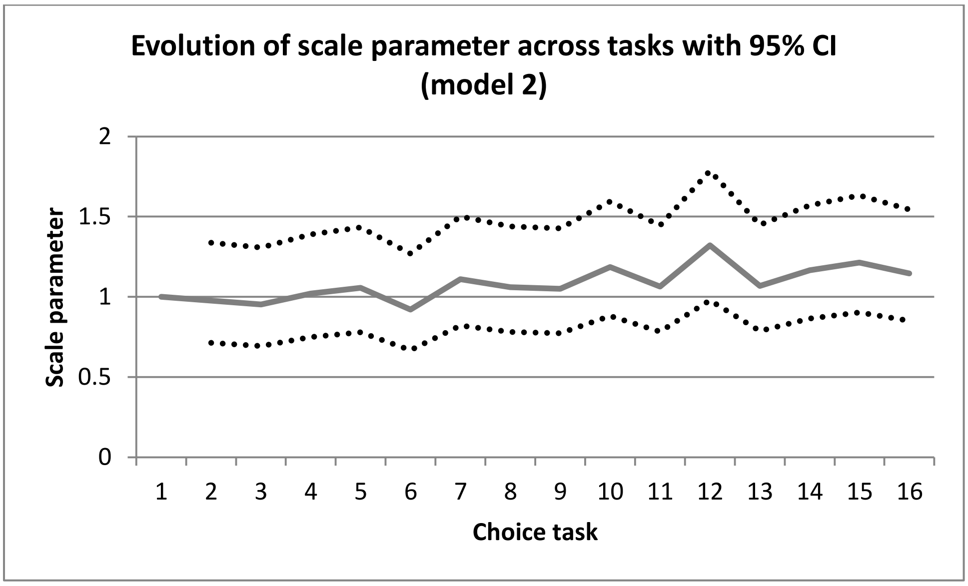
Table 2. Summary of model fit statistics for base models

	Log-likelihood	par.	LR test statistic against base model	Degrees of freedom	p-value
1. Base model	-3,713.76	7	-	-	-
2. Separate scale parameters	-3,703.54	22	20.45	15	0.156
3. Separate est. for all parameters	-3,651.66	112	124.20	105	0.097
4. Separate estimates for *manual*	-3,700.94	22	25.64	15	0.042
5. Separate estimates for *mech*	-3,703.55	22	20.42	15	0.156
6. Separate estimates for *burn*	-3,709.32	22	8.88	15	0.884
7. Separate estimates for *area*	-3,704.68	22	18.16	15	0.254

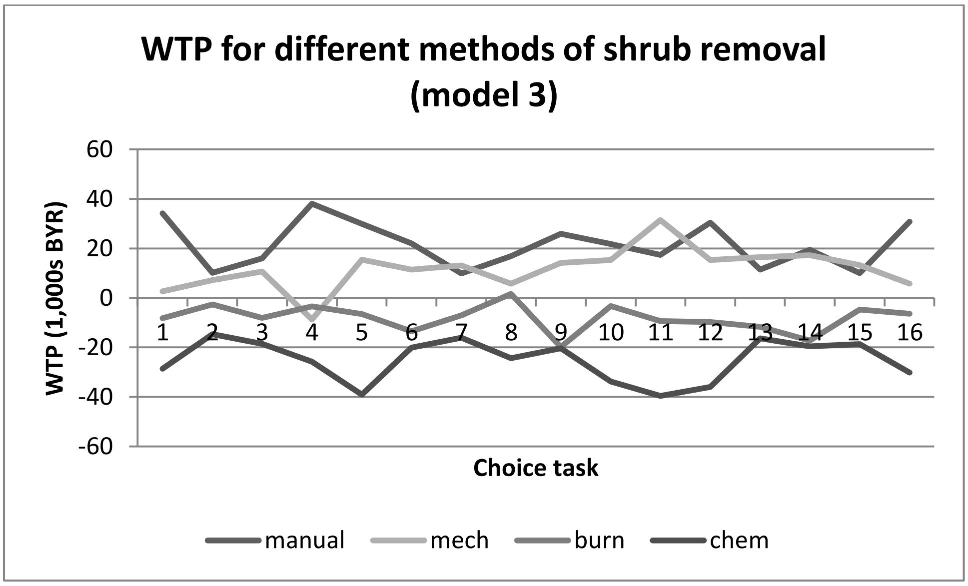
[8] Remembering the effects coding for chemical removal

8. Separate estimates for *reserve*	-3,707.21	22	13.10	15	0.594
9. Separate estimates for *cost*	-3,702.56	22	22.40	15	0.098
10. Separate estimates for *SQ*	-3,710.42	22	6.67	15	0.966

The insights from model 2 and 3 are further summarised in Figure 1. While we see fluctuations in scale and some evidence of an overall increase, the estimates from each stage fall within the 95% confidence interval of any of the estimates, and this is also in line with the observation that the gains obtained by model 2 were not significant. Looking at the various WTP measures coming out of model 3[9], we again see some fluctuation where no clear overall trend is visible.



[9] Noting that we also show a WTP for chem now given the effects coding.



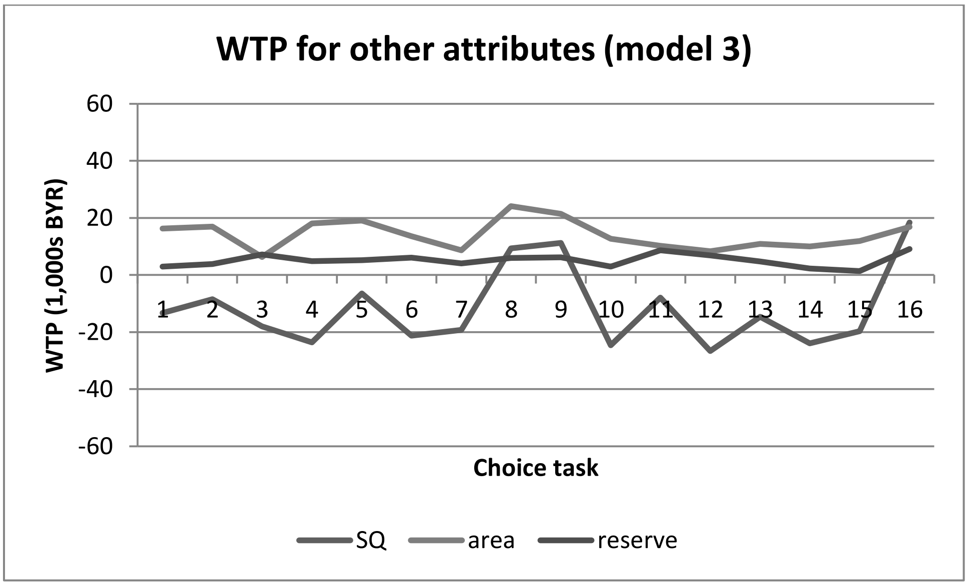


Figure 1: results across different choice tasks

4.2. Analysis with deterministic and random taste heterogeneity

Three models were estimated for the main analysis. We begin with a basic MNL model, with no random taste heterogeneity. This is then followed by a model allowing for random taste heterogeneity, using a specification with inter-respondent heterogeneity and intra-respondent homogeneity. In the final model, we allow jointly for inter-respondent and intra-respondent variation in tastes.

4.2.1. Model specification

The utility for the status quo alternative is given by a constant, i.e.

$$U_{1,n,t} = \delta_{SQ} + \varepsilon_{1,n,t} \quad , \tag{7}$$

where $\varepsilon_{1,n,t}$ is a type I extreme value term, distributed identically and independently across respondents and choices. In all our models, δ_{SQ} is kept fixed across respondents as well as across observations for the same respondent in the absence of any findings related to variations across respondents in the SQ constant after varying all remaining parameters.

The utility function for the three programme alternatives includes effects-coded parameters associated with the shrub removal method and continuous coefficients for the remaining three variables. Using the most general specification, we have:

$$\begin{aligned} U_{j,n,t} = {} & \left(\mu_{ma} + \sigma_{\alpha,11}\xi_{\alpha,1} + \sigma_{\gamma,1}\xi_{\gamma,1}\right)x_{eff-manual,j,n,t} \\ & + \left(\mu_{me} + \sigma_{\alpha,21}\xi_{\alpha,1} + \sigma_{\alpha,22}\xi_{\alpha,2} + \sigma_{\gamma,2}\xi_{\gamma,2}\right)x_{eff-mechanical,j,n,t} \\ & + \left(\mu_{bu} + \sigma_{\alpha,31}\xi_{\alpha,1} + \sigma_{\alpha,32}\xi_{\alpha,2} + \sigma_{\alpha,33}\xi_{\alpha,3} + \sigma_{\gamma,3}\xi_{\gamma,3}\right)x_{eff-burn,j,n,t} \\ & + \big(\mu_{ar} + \sigma_{\alpha,41}\xi_{\alpha,1} + \sigma_{\alpha,42}\xi_{\alpha,2} + \sigma_{\alpha,43}\xi_{\alpha,3} + \sigma_{\alpha,44}\xi_{\alpha,4} + \sigma_{\gamma,4}\xi_{\gamma,4} \\ & \qquad + \beta_{ar-user}z_{user,n}\big)x_{area,j,n,t} \\ & + \big(\mu_{re} + \sigma_{\alpha,51}\xi_{\alpha,1} + \sigma_{\alpha,52}\xi_{\alpha,2} + \sigma_{\alpha,53}\xi_{\alpha,3} + \sigma_{\alpha,54}\xi_{\alpha,4} + \sigma_{\alpha,55}\xi_{\alpha,5} + \sigma_{\gamma,5}\xi_{\gamma,5} \\ & \qquad + \beta_{re-educ}z_{educ,n} + \beta_{re-user}z_{user,n}\big)x_{reserve,j,n,t} \\ & + \left(\kappa_{cost} - e^{\mu_{co}+\sigma_{\alpha,co}\xi_{\alpha,6}+\sigma_{\gamma,co}\xi_{\gamma,6}} + {}^{\beta_{co-inc}}\!/_{inc_n}\right)x_{cost,j,n,t} \end{aligned}$$

$+\varepsilon_{j,n,t}$

where j=2,..,4.

With this notation, $\xi_{\alpha,1}$ to $\xi_{\alpha,6}$ represent standard Normal error terms that are distributed across respondents but not across choices for the same respondent, while $\xi_{\gamma,1}$ to $\xi_{\gamma,6}$ represent standard Normal error terms that are distributed across all observations[10].

We found heterogeneity in in the cost sensitivity to be highly significant and thus moved away from the relatively common approach in environmental valuation to keep the cost coefficient fixed. We also avoided relying on the Normal distribution for the cost coefficient as this would have led to undefined moments for the resulting WTP distributions (cf. Daly et al., 2012), using a negative Lognormal distribution instead[11].

Finally, four additional parameters were included to capture socio-demographic interactions, namely $\beta_{ar-user}$, which captures the shift in sensitivity to removal area size for users

---

[10] The non-cost coefficients are assumed to follow a Normal distribution, with correlation between the inter-respondent components but independently distributed intra-respondent components. The mean sensitivities are given by the μ parameters. For the inter-respondent heterogeneity, $\sigma_{\alpha,kl}, with\ 1 \leq l \leq k \leq 5$ represent Cholesky terms, with the variance for coefficient m being $\left(\sum_{l=1}^{m} \sigma_{\alpha,ml}{}^2\right)$ and the covariance between coefficients m and p (with m<p) being $\left(\sum_{l=1}^{m} \sigma_{\alpha,ml}\sigma_{\alpha,pl}\right)$. For the intra-respondent heterogeneity, $\sigma_{\gamma,k}, with\ 1 \leq k \leq 5$, represent the standard deviation estimates.

[11] We allowed for a non-zero bound through estimating an offset parameter κ_{cost}, in addition to the mean for the underlying Normal distribution of μ_{co} and standard deviation terms for the inter-respondent and intra-respondent components of $\sigma_{\alpha,co}$ and $\sigma_{\gamma,co}$ respectively. The cost coefficient was uncorrelated with other coefficients after initial modelling results showed a lack of correlation

(visitors) of the wetland reserve ($z_{user,n}$=1), $\beta_{re-educ}$ captures the shift in sensitivity to reserve size for respondents with a university degree ($z_{educ,n}$), $\beta_{re-user}$ captures the shift in sensitivity to reserve size for users of the reserve, and β_{co-inc} captures cost sensitivity as a function of income, with inc_n giving the income of respondent *n*.

All models were coded and estimated[12] in Ox 6.2 (Doornik, 2001). The standard errors reported in the tables take into account the repeated choice nature of the data in all three models[13]. We made use of MLHS draws (Hess et al., 2006) in estimation, with 500 inter-respondent draws per individual and per attribute, and 500 intra-respondent draws per observation and per attribute.

4.2.2. Model results

The modelling results are presented in Table 3. For all three models the signs of the main coefficients are the same and are consistent with *a priori* expectations. The estimate for the *SQ* constant is negative, indicating that respondents on average would like to move from the

---

[12] In the simple MNL model, we estimate the μ parameters as marginal sensitivities for the non-cost attributes, κ_cost as the cost coefficient, and finally the four interactions with socio-demographic characteristics. In the MMNL model with inter-respondent heterogeneity only, we additionally estimate μ_co and the sixteen σ_(α,·) parameters, while, in the MMNL model with additional intra-respondent heterogeneity, we also estimate the six σ_(γ,·) parameters.

[13] This is made possible by clustering together observations from the same respondent when calculating the BHHH matrix that enters the computation of the sandwich matrix. The latter is often referred to as giving robust standard errors – this is independent of whether observations for the same respondent are clustered together or not, doing so simply means that the computation also accounts for the repeated choice nature of the data, generally resulting in an upwards correction of standard errors (cf. Daly & Hess, 2011).

current situation to a programme of active protection. The positive and statistically significant estimates for the fixed MNL coefficients and the MMNL means for both *area* (μ_{ar}) and *reserve* (μ_{re}) imply that protection programmes associated with larger areas of shrub removal and the enlargement of the existing reserve are more likely to be chosen. Positive and statistically significant estimates for *manual* scything (μ_{ma}) and *mechanical* mowing (μ_{me}) indicate that people, on average, associate positive utility with these two methods, whereas controlled *burning* (μ_{bu}) and *chemical* herbicide use ($\mu_{ce} = -\mu_{ma} - \mu_{me} - \mu_{bu}$) contribute, on average, negatively to their utility. The MNL model shows negative and significant cost sensitivity (κ_{cost}), while the negative estimate for κ_{cost} in the MMNL models shows that the entire distribution for the cost sensitivity is negative (given the negative lognormal distribution for the remainder of the distribution).

Respondents with a higher income have lower price sensitivity (the impact of β_{co-inc} reduces with income), although this effect is statistically significant only in the MNL model. Respondents with a university degree have higher marginal utility associated with enlarging the reserve, while users, i.e. respondents who visited the wetland, have higher marginal utility for increases in the shrub removal area and increases in the reserve area, although the former is not statistically significant in the most complex model.

Table 3. Estimation results for three final models

			MNL		MMNL		MMNL + Intra	
			Coeff.	Asy. t-rat.	Coeff.	Asy. t-rat.	Coeff.	Asy. t-rat.
Mean effects	manual	μ_{ma}	0.4996	13.68	0.7272	8.40	0.8578	6.78
	mechanical	μ_{me}	0.2908	7.15	0.5763	7.30	0.6908	5.76
	burning	μ_{bu}	-0.1867	-4.56	-0.1537	-2.66	-0.1743	-2.05
	area	μ_{ar}	0.2814	12.04	0.4625	8.99	0.6620	7.47
	reserve	μ_{re}	0.076	6.68	0.1234	6.46	0.1588	5.95

	cost	μ_{co}		-4.1050	-21.62	-3.7854	-19.53
		κ_{cost} -0.0146	-12.54	-0.0044	-1.69	-0.0040	-1.28
	SQ constant	δ_{SQ} -0.4162	-4.68	-1.3619	-10.04	-1.2085	-7.01
Socio-demographic effects	Income	β_{co-inc} -0.0443	-10.23	-0.0140	-1.59	-0.0105	-1.07
	University educ.	$\beta_{re-educ}$ 0.0749	2.63	0.1903	3.60	0.2020	3.27
	User	$\beta_{re-user}$ 0.1942	7.24	0.1790	4.05	0.2086	3.81
		$\beta_{ar-user}$ 0.2848	6.42	0.3141	3.07	0.2187	1.73
Cholesky parameters (inter-respondent heterogeneity)		$\sigma_{\alpha,11}$		-1.0199	-12.04	-1.1792	-9.02
		$\sigma_{\alpha,21}$		0.1466	1.43	0.1760	1.64
		$\sigma_{\alpha,22}$		-0.8343	-9.85	-0.9851	-8.30
		$\sigma_{\alpha,31}$		0.2842	4.19	0.3104	3.78
		$\sigma_{\alpha,32}$		0.1925	2.90	0.2868	3.16
		$\sigma_{\alpha,33}$		-0.0432	-0.38	0.0671	0.88
		$\sigma_{\alpha,41}$		0.1798	4.27	0.2497	4.24
		$\sigma_{\alpha,42}$		-0.0828	-1.55	-0.0156	-0.27
		$\sigma_{\alpha,43}$		0.1622	2.02	0.6198	9.39
		$\sigma_{\alpha,44}$		0.5007	11.99	-0.0143	-0.23
		$\sigma_{\alpha,51}$		0.0433	2.18	0.0475	2.10
		$\sigma_{\alpha,52}$		-0.0019	-0.09	-0.0102	-0.46
		$\sigma_{\alpha,53}$		0.0686	2.44	0.0876	3.53
		$\sigma_{\alpha,54}$		0.0480	2.39	0.1484	5.51
		$\sigma_{\alpha,55}$		0.1061	3.67	0.0200	0.54
Inter-respondent cost heterogeneity		$\sigma_{\alpha,co}$		1.9895	13.73	1.7122	15.68
Intra-respondent heterogeneity parameters	manual	$\sigma_{\gamma,1}$				0.4398	1.46
	mechanical	$\sigma_{\gamma,2}$				0.6951	2.79
	burning	$\sigma_{\gamma,3}$				0.9944	3.92
	Area	$\sigma_{\gamma,4}$				0.0274	0.31
	Reserve	$\sigma_{\gamma,5}$				0.0080	0.18
	Cost	$\sigma_{\gamma,co}$				0.0100	0.15
	LL(β)	-3,570.4		-2,756.8		-2,747.7	
	Parameters	11		28		34	

The MMNL model with inter-respondent heterogeneity uses 17 additional parameters compared to the MNL model. We obtain an improvement in log-likelihood by 813.6 units, which is significant at high levels of confidence using a LR test[14]. The two parameters relating to the distribution of the cost coefficient are highly significant, while κ_{cost} remains negative, indicating an upper limit of the distribution below zero. Similarly, the means of the normally distributed parameters are all statistically significant as are the majority of the elements of the Cholesky matrix.

The MMNL model with additional intra-respondent heterogeneity has a further six parameters, relating to the standard deviations of the intra-respondent component, which are distributed independently from one another and across choices. This model obtains a further improvement in log-likelihood by 9.1 units for these six additional parameters, which gives us a LR-test p-value of 0.006 (again using a simple LR test). The improvements are much smaller than when moving from the MNL model to the first MMNL model, in common with past findings in the transport literature. In this model, the diagonal elements of the Cholesky matrix relating to inter-respondent heterogeneity for *reserve* and *cost* become insignificant, but this does not imply an absence of inter-respondent heterogeneity for these coefficients given the role of the off-diagonal elements. A closer inspection of the parameters relating to intra-respondent heterogeneity show that the gains in fit are the

[14] We use here a simple χ^2 critical value, where, with the null hypothesis being comfortably rejected, the issue of critical values being too high in the naïve approach are of little concern, noting only that with mixture χ^2 critical values, the null would be rejected even more strongly (see Chen & Cosslett, 1998; Moeltner & Layton, 2002).

result of capturing the intra-respondent variation for only three coefficients, namely those relating to the method of shrub removal (noting also the lower significance for $\sigma_{\gamma,1}$).

As a next step, Table 4 shows the implied heterogeneity patterns for each random coefficient, where we include the coefficient of variation (cv) to look at relative heterogeneity across models[15].

Table 4: Levels of inter-respondent and intra-respondent heterogeneity

	Model with inter-respondent heterogeneity only		Model with inter-respondent and intra-respondent heterogeneity					
			Inter-respondent heterogeneity		Intra-respondent heterogeneity		Combined heterogeneity	
	std. dev.	cv.	std. dev.	cv.	std. dev.	cv.	std. dev.	cv.
β_{ma}	1.02	1.40	1.18	1.37	0.44	0.51	1.26	1.47
β_{me}	0.85	1.47	1.00	1.45	0.70	1.01	1.22	1.77
β_{bu}	0.35	-2.25	0.43	-2.46	0.99	-5.71	1.08	-6.19
β_{ar}	0.56	1.22	0.67	1.01	0.03	0.04	0.67	1.01
β_{re}	0.14	1.15	0.18	1.13	0.01	0.05	0.18	1.14
$ln(-\beta_{co})$*	1.99	-0.48	1.71	-0.45	0.01	0.00	1.71	-0.45

* values relate the random component only, i.e. ignoring insignificant estimate of κ_{cost}

When moving from the model with inter-respondent heterogeneity only to the model with both inter-respondent and intra-respondent heterogeneity, we see small reductions in inter-

[15] The calculations relate only to the random part, excluding socio-demographic effects. We use the estimates from the Cholesky matrix to compute standard deviations for inter-respondent heterogeneity for the five non-monetary parameters For the intra-respondent heterogeneity, the standard deviations are obtained directly from the estimates of $\sigma_{\gamma,\cdot}$, while the taste heterogeneity estimates for cost relate to the underlying Normal distribution, i.e. the distribution of $ln(-\beta_{co})$, ignoring the insignificant offset parameter. The levels of heterogeneity for the combined intra-respondent and inter-respondent components are obtained through simulation.

respondent heterogeneity for all coefficients except β_{bu}, despite the fact that intra-respondent heterogeneity was only retrieved for the first three parameters. The level of intra-respondent heterogeneity is lower than the level of inter-respondent heterogeneity for β_{ma} and β_{me}, but is substantially higher for intra-level heterogeneity for β_{bu}. The possible explanation of this phenomenon is that on average, respondents' preferences with respect to *burning* are not well defined. Although *controlled burning* was used extensively in the past, current restrictions have reduced its use, making respondents relatively unfamiliar with the method, while anti-burning actions carried out by local media in the recent years may also contribute to respondents' uncertainty. When comparing the total levels of heterogeneity in the two models, we can see an increase in the heterogeneity for the three coefficients relating to method of removal, with a small reduction in the heterogeneity for the area, reserve and cost coefficients.

Table 5 contains the calculated correlations between the five normally distributed non-cost coefficients. We obtain similar patterns of correlation for the two MMNL models, with strong negative correlation between *manual* and *burning*, which makes sense with the two methods being very different. We also see relatively high positive correlation between *area* and *reserve*, again consistent with intuition. Indeed, the larger the area from which the shrubs are removed, the better are the conditions for the rare bird species. Similarly the larger the reserve is, the better the conservation of the site is. The main difference between the two models comes in the increased correlation between *area* and *burning* in the second MMNL model, which could possibly be explained on the grounds that respondents who value very large areas of shrub removal see *burning* as an efficient method.

Table 5: correlations between normally distributed non-cost coefficients

Model with inter-respondent heterogeneity only						Model with inter-respondent and intra-respondent heterogeneity				
	β_{ma}	β_{me}	β_{bu}	β_{ar}			β_{ma}	β_{me}	β_{bu}	β_{ar}
β_{me}	-0.17					β_{me}	-0.18			
β_{bu}	-0.82	-0.41				β_{bu}	-0.73	-0.53		
β_{ar}	-0.32	0.20	0.14			β_{ar}	-0.37	0.09	0.40	
β_{re}	-0.31	0.07	0.18	0.54		β_{re}	-0.26	0.10	0.23	0.53

As a final step, we look at the WTPs calculated from the model estimates. The calculated trade-offs are reported in Table 6. The values are calculated for each person in the data, taking into account the socio-demographic interactions, and hence we also obtain heterogeneity in the MNL model. Starting with the MNL estimates, *manual* removal is valued more highly than *mechanical* removal, and *burning* is valued less negatively than *chemical* removal. The WTP for area of removal and reserve size are positive, with the former being about twice as large. This indicates that respondents prefer active protection to simply enlarging the reserve.

Table 6. Implied WTP measures (in thousands of BYR)

	MNL			MMNL			MMNL + Intra		
	mean	std. dev.	c.v.	mean	std. dev.	c.v.	mean	std. dev.	c.v.
manual	21.83	5.04	0.23	38.27	77.59	2.03	41.64	90.88	2.18
mechanical	12.71	2.93	0.23	30.34	63.94	2.11	31.53	83.31	2.64
burning	-8.16	1.88	-0.23	-8.07	24.86	-3.08	-8.76	69.32	-7.91
chemical	-26.38	6.09	-0.23	-60.54	79.17	-1.31	-64.41	119.79	-1.86
Area	14.54	6.07	0.42	27.36	46.06	1.68	33.43	52.84	1.58

Reserve	3.32	0.77	0.23	6.50	11.24	1.73	7.64	13.48	1.76

Moving to the MMNL model with inter-personal variation only, we see increases in the mean WTP measures for the majority of components, and substantial levels of heterogeneity across respondents. The ordering of WTP remains the same, but the gap between area and reserve increases substantially. This is further increased when moving to the model allowing for intra-respondent heterogeneity, where we see increases in mean values as well as standard deviations. This applies to all attributes, reflecting the fact that a failure to account for intra-respondent heterogeneity can affect parameters other than the one in which the heterogeneity occurs. The relative levels of heterogeneity increase for the WTP for the four methods of shrub removal, especially for *burning*, while they decrease slightly for the WTP for *area* and increase slightly for the WTP for *reserve*.

4.2.3. Initial insights into relationship between uncertainty in CV and intra-respondent heterogeneity in SC

As a final step in this paper, we provide a comparison of the results from the CV question and the models estimated on the SC scenarios. This investigation is not meant to provide a full consistency check between the two data sources, but merely initial insights into the relationship between uncertainty in CV and intra-respondent heterogeneity in SC. The CV valuations in Table 7 show the mean and median valuation across respondents of the midpoints of the reported ranges, the standard deviation of that midpoint valuation across respondents, and the standard deviation of the valuations when factoring in the uncertainty at the individual level (which is zero for respondents without uncertainty). This latter measure is obtained by assuming that the valuations at the individual respondent level follow a uniform distribution between the lower and upper limits of the reported ranges.

Table 7. WTP estimates from CV exercise and equivalent SC valuations (thousands of BYR)

	CV	**MNL**	**MMNL**	**MMNL + Intra**
mean valuation	54.41	107.88	174.75	185.19
median valuation	30	112.78	105.9	105.85
std dev of valuation (inter-respondent)	63.27	27.12	196.69	-
std dev of valuation (inter- and intra-respondent)	65.7	-	-	222.16
cv. of valuation (inter-respondent)	1.163	0.251	1.126	-
cv. of valuation (inter- and intra-respondent)	1.208	-	-	1.199

In order to be able to compare estimates from the CV exercise with the SC results, we calculate for the latter the Hicksian welfare measure for a change from SQ to the programme of 2 000 ha/year of mechanical mowing and 4 000 ha of the reserve enlargement which is based on the compensating variation log-sum formula described by Hanemann (1984)[16]. This gives us higher mean and median values than for the CV, but this is to be expected given the evidence from contingent valuation studies that mean WTP estimated using the dichotomous choice question format in most studies substantially exceeds mean WTP obtained using the open-ended question format (Walsh et al. 1992;

[16] We acknowledge a degree of simplification in this comparison, not least as we're combining evidence from multiple parameters estimated on the SC data, each with their own estimation error, and compare this to a single measure from CV.

Johnson et al. 1990; Schulze et al. 1996)[17]. In addition to that, most studies which compare SC estimates with DC-CV estimates report the former to be substantially larger (e.g. Hanley et al. 1998; Barrett et al. 1996; Stevens et al. 2000). We note that the difference in means become larger as we move away from the MNL model, which was to be expected given the use of the lognormal distribution for the cost coefficient. Additionally, the calculations above assume that the endpoints in the CV ranges are bounds of the distribution rather than say 95% confidence limits. The median valuations are rather consistent across the three choice models. The coefficient of variation is quite similar between the MMNL model (1.126) and the CV results (1.163), but obviously much lower in the MNL model. In the CV results, we see only a small increase in the coefficient of variation[18], by 3.8% (from 1.163 to 1.208) when factoring in intra-respondent variation. In the SC data, the coefficient of variation in the model with both layers of heterogeneity is remarkably similar to that from the CV values (1.199 *vs* 1.208), suggesting greater consistency between the two when factoring in intra-respondent heterogeneity, where this gives a 6.5% increase in heterogneity. The fact that the share of overall heterogeneity due to intra-respondent variations seems lower in the CV measures than in the WTP measures coming from the SC data is potentially not surprising as there may be more sources for intra-respondent variation in SC than CV. Nevertheless, the consistency in results is remarkable.

---

[17] SC can be considered as a variant of DC-CV whereas CIOE question is a variant of open-ended question, so these findings are relevant to our study.

[18] We work here with the coefficient of variation given the differences in mean levels between CV and SC but also the small changes in mean levels we see in SC when incorporating intra-respondent heterogeneity.

5. Summary and conclusions

To the best of our knowledge, all previous applications of MMNL in environmental valuation have assumed homogenous sensitivities across choices for the same respondent, an assumption which we have questioned in this paper in the context of a SC survey on wetland conservation in Belarus.

After an exploration of generic (across respondents) variations at the MNL level across tasks in both scale and relative sensitivities revealed no clear trends, we turned our attention to a random treatment of such heterogeneity across tasks within a MMNL model allowing for both inter- and intra-respondent variations. In this model, the actual level of within respondent heterogeneity is held constant across respondents, but the direction of the variation can vary across respondents. In other words, for some respondents, a given sensitivity might increase as we go through the tasks, while it might decrease for other respondents, or fluctuate randomly for yet another group. The assumption of constant degrees of intra-respondent heterogeneity across respondents is a simplification and that the levels of heterogeneity, especially if caused by uncertainty, may vary across respondents. Such a further development is theoretically possible but would lead to further big increases in identification and estimation complexity.

We found that the model with additional intra-respondent heterogeneity significantly improves model fit. A further inspection showed that the gains in fit are the result of capturing the intra-respondent variation for three coefficients out of six, namely those relating to the method of shrub removal. These attributes were identified during the focus groups as potentially the least familiar to respondents. We believe that this may increase the scope for intra-respondent heterogeneity, especially for the controversial *burning* approach, where the level of intra-respondent variation was found to be substantially higher than the inter-respondent heterogeneity[19].

It is also worth noting that our discussions thus far have focussed on the interpretation of random parameters relating solely to heterogeneity in sensitivities. In the case of categorical attributes such as for the method of shrub removal, it should be noted that random parameters on these terms can also be interpreted as error components that capture heteroscedasticity across the alternatives in a quasi-labelled type of choice scenario. The fact that intra-respondent variation is then only retrieved for these terms could suggest that the intra-respondent heterogeneity picks up variations in the structural part of the utility, but also introduces a more flexible error structure (or at least partly picks up these effects).

We also tested for the impact on WTP measures of not accounting for intra-respondent heterogeneity, where our result show impacts on inter-personal mean and variations of WTP measures. Importantly, our results show that a failure to account for intra-respondent heterogeneity can affect parameters other than only the ones in which heterogeneity occurs.

[19] Unlike with mechanical mowing or chemical process, people were familiar with burning, and during the focus groups, respondents had mixed feelings about the possibility of burning being used. In our opinion, there are two possible explanations for the large intra-respondent variation for burning. The first is that the tradition of using burning in the past, combined with anti-burning actions carried out in Belarus in recent years, could result in respondents' uncertainty regarding the use of this method for an active protection. The other possible explanation is that most people were familiar with burning but only on a small scale. The use of this method on such a large scale (1000-4000 ha) was novel to the respondents during the focus group.

As a further novel angle, we provided some initial insights into the possible link between our findings and discussions in the CV literature on respondent uncertainty. According to DPH and CA, preference dynamics are a result of learning. Increased experience should therefore reduce the obtained WTP-ranges over the choice sequence. Our model does not allow us to test these dynamics as the variation is constant across tasks, and so we are not able to test whether WTP-ranges decrease over the choice sequence. In our model, we link the random variations in the relative sensitivities to the WTP-ranges observed in CV studies and interpret this as uncertainty, motivated in part by an absence of significant trends in the variations. We find surprisingly strong consistency of results between the two formats in terms of the relative level of heterogeneity albeit that we note a slightly bigger role for intra-respondent variations in the SC values. This is not surprising as the scope for intra-respondent heterogeneity is arguably greater in SC given the within respondent variations in the scenarios presented to them, while, for the present CV data, all intra-respondent heterogeneity relates solely to uncertainty in a single setting, i.e. an absence of dynamic effects. Nevertheless, the similarity in the levels of heterogeneity does lend some empirical support to the notion that the majority of intra-respondent heterogeneity in models estimated on SC data may relate to uncertainty, when no clear dynamic effects have been identified.

While the extent of intra-respondent variation and the gains in fit are small overall, compared to the introduction of inter-respondent heterogeneity, we feel that the method has relevance also in environmental economics, where, in many applications of SC, people are unfamiliar with valued goods/services. In such situations it is difficult to expect that respondents have a-priori well-formed preferences and we believe that this creates substantial scope for intra-respondent variation of a type that maybe also cannot easily be accommodated in a deterministic manner. Additionally, the estimation of a model allowing for intra-respondent heterogeneity can highlight the existence of such variations in sensitivities and motivate an analyst to find the drivers thereof. Of course, it is important to keep in mind that any gains in insights need to be traded off with the associated increase in computational cost. Finally, the identification of intra-respondent heterogeneity is clearly a data hungry process, and this may motivate the use of full preference orderings rather than data on first choices alone, but our recent work with best worst data (Giergiczny et al., 2013) has made us wary of such data augmentation approaches.

ACKNOWLEDGEMENTS

The authors are grateful for the comments of Thijs Dekker and one anonymous referee which helped to significantly improve the paper. The authors would like to thank Sviataslau Valasiuk for providing the SC and CV data. The work on this paper was supported by the Polish National Science Centre grant DEC-2012/07/E/HS4/04037.

REFERENCES

Adamowicz, W., Boxall, P., Williams, W., and Louviere, J.J. (1998), Stated preference approaches for measuring Passive use values: choice experiments and contingent valuation Amer. J. Agr. Econ. 80, 64-75.

Ariely, D., Loewenstein, G., Prelec, D. (2006), Tom Sawyer and the construction of value, Journal of Economic Behavior & Organization, vol. 60, no. 1, pp. 1-10.

Ariely, D., Loewenstein, G., Prelec D. (2003), Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences, Quarterly Journal of Economics 118 (1): 73–105.

Baerenklau, Kenneth A. and Bill Provencher. "Static modeling of dynamic recreation behavior: implications for prediction and welfare estimation", Journal of Environmental Economics and Management , 50(3): 617-636 (2005).

Barrett, C., Stevens, T.H. and Willis, C, (1996), Comparison of CV and conjoint analysis in groundwater valuation. In: Herriges, J. (Compiler), Ninth Interim Report, W-133 Benefits and Costs Transfer in Natural Resource Planning, Department of Economics, Iowa State University, Ames, IA.

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

Bateman, I.J., Langford, I.H., Jones, A.P., Kerr, G.N. (2001), Bound and path effects in multiple-bound dichotomous choice contingent valuation, Resource and Energy Economics, 23(3): 191-213.

Bhat, C. R. & Castelar, S. (2002), A unified mixed logit framework for modelling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. Transportation Research Part B 36 (7), 593-616.

Bhat, C. R. & Sardesai, R. (2006), The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B 40 (9), 709-730.

Bradley, M and Daly, A. (1994), Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data, Transportation, 21, 2, 167-184

Brouwer, R., Dekker, T., Rolfe, J., Windle J. (2010), Choice Certainty and Consistency in Repeated Choice Experiments, Environmental and Resource Economics, vol. 46, no. 1, pp. 93-109.

Cameron, T.A. & Quiggin, J. (1994), Estimation using contingent valuation data from a “dichotomous choice with follow-up” questionnaire, Journal of Environmental Economics and Management, 27, 218-234.

Chen, H.Z. and Cosslett, S.R. (1998), Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach, American Journal of Agricultural Economics, 512-520.

Cherchi, E., Cirillo, C., Ortúzar, J. de D. (2009), A mixed logit mode choice model for panel data: accounting for different correlation over time periods, paper presented at the International Choice Modelling Conference, Harrogate.

Daly, A.J. & Hess, S. (2011), Simple Approaches for Random Utility Modelling with Panel Data, paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C.

Daly, A.J., Hess, S. & Train, K.E. (2012), Assuring finite moments for willingness to pay in random coefficients models, Transportation 39(1), pp. 19-31.

Dekker, T., Hess, S. Brouwer R., and Hofkes H. (2013), Implicitly or explicitly uncertain? International Choice Modelling Conference, Sydney.

Dekker, T. (2012), Preference uncertainty in stated choice experiments, doctoral thesis.

Doornik, J.A. (2001), Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

Ellingson L, Boman M, Norman J, Mattson L. (2009), Valuation uncertainty and maximum willingness to pay: A nonparametric comparison between elicitation formats. Southern Swedish Forest Research Centre, Alnarp.

Giergiczny M., Hess S., Dekker T., Chintakayala P.K. (2013), Testing the consistency (or lack thereof) between choices in best-worst surveys. paper presented at the Third International Choice Modelling Conference, Sydney.

Hanemann, W. (1984), Welfare Evaluations in Contingent Valuation Experiment with Discrete Responses', American Journal of Agricultural Economics, Vol. 66, pp. 332-341.

Hanley, N., Kriström B., Shogren J.F. (2009), Coherent Arbitrariness: On Value Uncertainty for Environmental Goods. Land Economics 85, 41-50.

Hanley, N., Adamowicz, W., Wright, R.E. (2005), Price vector effects in choice experiments: an empirical test, Resource and Energy Economics, vol. 27, no. 3, pp. 227-234.

Hanley, N., Wright, R., Adamowicz, W. (1998), Using choice experiments to value the environment, Environmental & Resource Economics, Vol. 11, pp. 413– 428.

Håkansson, C. (2008), A new valuation question: Analysis of and insights from interval open ended data in contingent valuation. Environ Resour Econ 39(2): 175-188.

Hensher, D., Rose, J. M., & Beck, M. (2011), Are there specific design elements of choice experiments and types of people that influence choice response certainty?, in 2nd International Choice Modelling Conference, Leeds, UK.

Hensher, D. A., Greene, W. H., (2003). The Mixed Logit Model: The State of Practice. Transportation 30 (2), 133-176.

Hess, S., Hensher, D. & Daly, A.J. (2012), Not bored yet – revisiting respondent fatigue in stated choice experiments, Transportation Research Part A 46(3), pp. 626-644.

Hess, S. & Rose, J.M. (2009), Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B 43(6), 708-719.

Hess, S. & Train, K.E. (2011), Recovery of inter- and intra-personal heterogeneity using mixed logit models, Transportation Research Part B, 45(7), pp. 973-990.

Hess, S., Train, K.E. & Polak, J.W. (2006), On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice, Transportation Research Part B, 40(2), pp. 147-163.

Holmes, T. P. & Boyle, K.J. (2005), Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions, Land Economics, vol. 81, no. 1, pp. 114-126.

Hoyos, D. (2010), The State of the Art of Environmental Valuation with Discrete Choice Experiments. Ecological Economics, Vol. 69, pp. 2372-2381.

Johnson, R., N. Bregenzer, and Shelby B. (1990), Contingent Valuation Question Formats: Dichotomous Choice versus Open-Ended Responses." In Economic Valuation ofNatural Resources: Issues, Theory, and Applications, eds., R. Johnson and G. Johnson, pp. 193-204. Boulder CO: Westview Press.

Kobayashi, M., Moeltner, K. and Rollins, K. (2012). Latent thresholds analysis of choice data under value uncertainty, American Journal of Agricultural Economics 94(1): 189-208.

Li C-H, Mattsson, L. (1995), Discrete choice under preference uncertainty: an improved structural model for contingent valuation. J Environ Econ Manage 28(2):256–269

Lundhede T. H., Olsen S. B., Jacobsen J. B. & Thorsen B. J. (2009), Handling respondent uncertainty in Choice Experiments: Evaluating recoding approaches against explicit modeling of uncertainty. Journal of Choice Modelling, 2(2), 118-147.

Mahieu, P. A., Riera, P., Giergiczny, M. (2012), The Influence of Cheap Talk on Willingness to Pay Range. Some Empirical Evidence. Journal of Environmental Planning and Management, 55 (6), 753-763.

McFadden, D. (1994), Contingent valuation and social choice, American Journal of Agricultural Economics, 76: 689-708.

McFadden, D., Train, K., (2000), Mixed MNL Models for discrete response. Journal of Applied Econometrics 15, 447-470.

Moeltner, K. and Layton, D. (2002). A censored random coefficients model for pooled survey data with application to the estimation of power outage costs, Review of Economics and Statistics 84(3): 552-561.

Norwood, F. B. (2005), Can Calibration Reconcile Stated and Observed Preferences? Journal of Agricultural and Applied Economics 37(1): 237-248.

Plott, C.R. (1996), Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis, in Rational Foundations of Economic Behavior. K. Arrow, E. Colombatto, M. Perleman, and C. Schmidt, eds. London: Macmillan and NY: St. Martin’s, pp. 225–50.

Revelt, D. & Train, K. (1998), Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4), pp. 647-657.

Rose, J.M. (2014), Interpreting discrete choice models based on Best-Worst data: A matter of framing, paper presented at the 93rd Annual Meeting of the Transportation Research Board.

Schulze, W., G. McClelland, D. Waldman, and J. Lazo. (1996), Sources of Bias in Contingent Valuation. In The Contingent Valuation of Environmental Resources, eds., D. Bjornstad and J. Kahn, pp. 97-116. Cheltenham, United Kingdom: Edward Elger,.

Slovic, P. (1995), The construction of preference, American Psychologist, vol. 50, no. 5, pp. 364-371.

Stevens, T. H., Belkner, R., Dennis, D., Kittredge, D., Willis, C. (2000), Comparison of contingent valuation and conjoint analysis in ecosystem management', Ecological Economics, Vol. 32, pp. 63-74

Street, D. J., & Burgess, L. (2007), The Construction of Optimal Stated Choice Experiments: Theory and Methods. Wiley-Interscience.

Street, D. J., Burgess, L., Louviere, J. J. (2005), Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing, 22(4):459–470.

Train, K. (2003), Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, MA.

Valasiuk, S., Giergiczny, M., Klimkowska, A., Żylicz, T. (2013), Are people in transitional economies willing to pay for ecosystem services and biodiversity? A choice experiment for wetland protection - the Case of Zvaniec Fen Mire in Belarus. Working Paper.

Walsh, R., D. Johnson, and J. McKean. (1992), Benefit Transfer of Outdoor Recreation Demand Studies: 1968-1988. Water Resour. Res. 28:707-13.

Welsh MP, Poe GL. (1998), Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. J Environ Econ Management 36:170–185

Yañez, M.F., Cherchi, E., Heydecker, B., Ortúzar, J. de D. (2011), On the Treatment of Repeated Observations in Panel Data: Efficiency of Mixed Logit Parameter Estimates. Networks and Spatial Economics, forthcoming.