University of St. Andrews

Discussion papers in Environmental Economics

http://www.st-andrews.ac.uk/gsd/research/envecon/eediscus/

Paper 2016-16

Linking perceived choice complexity with scale heterogeneity in discrete choice experiments: home heating in Finland

E. Ruokamo, M. Czajkowski, N. Hanley, A. Juutinen, and R. Svento

Keywords: Home heating, Choice experiment, Choice modelling, Scale heterogeneity, Generalized mixed logit, Complexity

JEL codes: D12, Q40, Q48, Q51, Q55

Ruokamo, E.[*] [1, 2], Czajkowski, M.[3], Hanley, N.[4], Juutinen, A.[2, 5] & Svento, R.[2]

[*] Corresponding author (e-mail: enni.ruokamo(at)oulu.fi)
[1] *Finnish Environment Institute (SYKE), Oulu, Finland*
[2] *Department of Economics, Oulu Business School, University of Oulu, Finland*
[3] *Department of Economic Sciences, University of Warsaw, Poland*
[4] *Department of Geography and Sustainable Development, University of St Andrews, Scotland, UK*
[5] *Natural Resources Institute Finland (Luke), Oulu, Finland*

Linking perceived choice complexity with scale heterogeneity in discrete choice experiments: home heating in Finland

Abstract

Choosing a specific heating system is a complex and difficult decision for homeowners as there exists a wide array of heating technologies with different characteristics that one can consider before purchasing. We include multiple heating technologies and attributes in our Choice Experiment design and explore the effect of perceived choice complexity on the randomness of choices. In particular, we investigate how different self-evaluated factors of choice complexity affect mean scale and scale variance. Our findings suggest that perceived choice complexity has a systematic impact on the parameters of econometric models of choice. However, there are differences between alternative self-evaluated complexity-related covariates. Results indicate that individuals who report that answering the choice tasks was difficult have less deterministic choices. Perceptions of the realism of home heating choice options also affect scale and scale variance.

JEL classification: D12, Q40, Q48, Q51, Q55

Keywords: Home heating, Choice experiment, Choice modelling, Scale heterogeneity, Generalized mixed logit, Complexity

1. Introduction

Decision making in Choice Experiments (CEs) involves respondents in comparing options described in terms of attributes and attribute levels, and making trade-offs between these attributes. According to random utility theory, individuals make choices between options based on the utility they obtain from the attributes used to describe these options, but with a degree of randomness (Thurstone 1927). The random component of the utility can be interpreted either as partly random choices from the perspective of the individual making that decision, due for example to preference uncertainty; or the random part can be due to the inability of the economist to measure everything that determines choices (Czajkowski et al. 2014b). It would seem likely that randomness from either perspective should be related to the complexity of the choice process. In this paper, therefore, the focus is on observing and measuring possible indicators of choice complexity, and then testing how these determine the randomness of choices and this randomness varies across consumers.

There is a wide literature focused on the determinants of choice consistency. This covers issues such as the use of choice heuristics, the level of care and attention that individual gives to a choice situation, and the effects of familiarity with the good on how random peoples' choices are (see e.g., Bush et al. 2009; Börger 2015; Campbell et al. 2008; Czajkowski et al. 2014b; Erdem et al. 2014; Hess and Stathopoulos 2013; LaRiviere et al. 2014; Scarpa et al. 2009). Task complexity and choice uncertainty (meaning how certain the respondents are about their choices) can also be sources of inconsistent choices (Beck et al. 2013; DeShazo and Fermo 2002; Lundhede et al. 2009). However, the effects of perceived choice complexity on the randomness of choice are currently under-explored.

Currently, the most flexible model which allows one to simultaneously control for unobserved preference and scale heterogeneity (even without allowing the preference parameters to be correlated) is the Generalized Mixed Logit model (G-MXL) (Fiebig et al. 2010; Greene and Hensher 2010; Keane and Wasi 2013). Czajkowski et al. (2016) show how to expand this model to introduce observable sources of mean scale and scale variance differences. In this paper, we use their framework to explore how different self-evaluated factors of choice complexity affect the scale parameter.

The scale parameter is a key behavioral factor in random utility choice models, as it weights the importance of the deterministic part of the utility relative to the random component. A failure to acknowledge scale heterogeneity in the modelling processes may induce biases in subsequent measures of willingness to pay (WTP) estimates and associated welfare analysis and policy recommendations. To investigate the extent of this bias, we also examine the differences in preference estimates between respondents who report higher levels of perceived choice complexity and those who report lower levels.

Fiebig et al. (2010) suggest that scale heterogeneity is rather more important in contexts involving complex choice objects. In this paper, we use data from a CE (Choice Experiment) survey where homeowners' hypothetical home heating system choices are recorded. Choosing a specific heating system is a complex and difficult decision for homeowners (Decker and Menrad 2016). In many countries, a heating system is a necessity for maintaining a suitable ambient temperature in one's home, and for supplying hot water. Difficulties rise from the facts that investments in a heating system are made only occasionally (for example, made only once every 20 years) and that the investment is a significant expenditure for most households. More importantly, there exists a wide array of heating technologies with different characteristics (e.g., price factors, comfort of use aspects, ecological and technical issues of the heating systems) that one can consider carefully before purchasing. We included multiple heating technologies and attributes in our CE design. As a result, the choice tasks were rather complex.

It is reasonable to speculate that some individuals found the choice tasks in our survey more difficult than others. We can hypothesize that higher levels of perceived complexity lead to a decrease in choice accuracy, i.e., an increase in choice randomness. This can be tested, provided one has measures of choice set complexity as perceived by the respondent. This study therefore explores the link between perceived choice complexity and choice randomness. We make a novel use of respondents' self-evaluated factors concerning choice complexity and test the effects of these on the estimated randomness in the choices made. To our knowledge, this way of explaining scale heterogeneity with self-evaluated complexity covariates has not been done before in CE studies.

The complexity covariates we are interested in capture different aspects of this broad concept. The examined covariates are the perceived difficulty of the choice tasks and the perceived unrealism of the choice scenarios. The econometric G-MXL model controls for the effect of these covariates on choice consistency by allowing the model's scale parameter to be a function of them. If the scale parameter increases, the deterministic part of the utility is assigned a greater weight compared to the error term. As a result, higher levels of scale indicate more deterministic (i.e., less random) choices.

Using this approach, we test the following hypotheses: (H1) individuals who report that answering the choice tasks was difficult have less deterministic choices than those who considered the choices to be easy, i.e., mean scale should decrease as perceived difficulty increases and (H2) individuals who report that there were unrealistic choice scenarios making answering complicated will have less deterministic choices, i.e., mean scale should decrease as perceived unrealism increases.

Our measures of choice complexity have significant effects on scale heterogeneity. Individuals who stated that the choice tasks were difficult have less deterministic choices. Additionally, if respondents report that the perceived unrealism of choice alternatives made answering complicated, they seem to have more random choices, and further, they seem to have more similar scale parameters. Perceived complexity, however, does not seem to induce significant biases in preference parameters in this dataset, and therefore has no significant effect on welfare estimates.

The reminder of this paper is organized as follows. Section 2 reviews the previous literature regarding both choice complexity and possible factors contributing to (variations in) scale. Section 3 shows how scale and preference heterogeneity is modelled. Then the case study and the complexity-related covariates are presented. Results are provided in Section 5. Finally, we draw some conclusions.

2. Literature review

Heiner (1983) argues that individuals cannot fully decipher the complexity of the situations they face and thus make seemingly sub-optimal decisions. He suggests that the complexity and uncertainty surrounding a choice situation often lead individuals to adopt simplified strategies, and that more effort should be expended to understand the role that complexity plays in choice behavior. Since then, many have explored diverse dimensions of choice complexity on decision making (see e.g., Boxall et al. 2009; DeShazo and Fermo 2002; Simonson and Tversky 1992; Regier et al. 2014; Swait and Adamowicz 2001a, 2001b).

It is widely acknowledged that task complexity increases with the number of alternatives and attributes used to describe the good (DeShazo and Fermo 2002; Hensher 2006; Louviere 2001; Swait and Adamowicz 2001a, 2001b). Swait and Adamowicz (2001a, 2001b) employ a concept of entropy to measure task complexity. Their entropy measure is simultaneously a function of number of alternatives, the number of attributes, the relationship between the attribute vectors themselves, and the structure of preferences, as an indicator of choice task complexity. Their findings suggest that a simpler processing strategy is used in cases with high levels of task complexity. DeShazo and Fermo (2002) present similar findings. They measure the effects of the quantity of information contained in the choice set by varying the number of alternatives as well as number of attributes in the choice tasks, and changing the correlation structures within and between alternatives. They report that all these measures of choice complexity affect choice consistency, and further, distort welfare estimates.

One potential factor affecting to choice complexity is the degree of confidence a respondent has in the choices they make. Previous studies that have examined scale as a function of choice certainty show that as respondent certainty about their choices decreases, their choices become less deterministic (Beck et al. 2013; Börger 2015; Lundhede et al. 2009).

Research focusing on factors contributing to scale heterogeneity and utilizing the G-MXL model is growing (Börger 2015; Christie and Gibbons 2011; Czajkowski et al. 2014a, 2014b, 2014c, 2016; Fiebig et al. 2010; Juutinen et al. 2012; LaRiviere et al. 2014). Christie and Gibbons (2011) argue that valuation of complex and unfamiliar goods requires a measure of whether respondents participating in such studies are able to construct and later reveal their true preferences. They suggest that it is particularly important to account for scale heterogeneity when individuals are required to choose between complex and unfamiliar goods and services, or when the choice task is cognitively challenging. Czajkowski et al. (2014b; 2016) investigate the effects of familiarity on choice consistency. In Czajkowski et al. (2014b), a model is constructed, which allows for individuals to learn about their preferences through consumption experience. Their main finding is that individual's scale increases and the variance of scale decreases with experience. Czajkowski et al. (2016) develop an approach for controlling the effects of different information sets provided to respondents. In particular, they allow information to affect preferences as well as the mean and variance of individual-specific scale parameters. Their findings indicate that the information set provided to respondent affects the mean of individuals' scale parameters and its variance, however, the preference parameters are not that sensitive to changes in information.

Czajkowski et al. (2014a) examine ordering effects in CEs. They demonstrate that respondents' learning and fatigue may lead to changes in preference parameters as well as the variance in its error terms (i.e., scale). To investigate scale dynamics they include choice task numbers as explanatory variables of scale. They observe that respondents' choices became more deterministic in the number of completed choice tasks, but do not find evidence on scale decrease while going further in the choice tasks. Besides ordering effects, there are also other factors that affect choice randomness. Börger (2015) investigates the effect of response time on scale. His results indicate that longer response time is associated with higher scale. Proxies for the cognitive abilities of individuals (for example education and age) have been added as covariates of scale (Czajkowski et al. 2014c; Juutinen et al. 2012), however, no systematic effects has been found so far.

3. Modelling approach

The CE technique is an application of the characteristics based theory of value (Lancaster 1966) combined with random utility theory (Thurstone 1927). According to random utility theory, individuals make choices based on the presence of the characteristics of the good in question with some degree of randomness. A frequently-used specification in choice modelling is the Mixed Logit (MXL) model which takes into account preference heterogeneity (see Ben-Akiva et al. 1997; McFadden 1974; Revelt and Train 1998; Train 2009). In the MXL model the utility of individual *n* choosing alternative *j* in the choice situation *t* is represented in the following general form

$$U_{njt} = \sigma \beta_n' x_{njt} + \varepsilon_{njt} . \tag{1}$$

where x_{njt} is a vector of non-cost and cost attributes, β_n is a vector of estimated parameters and ε_{njt} is an idiosyncratic error. Note that in our case study, x_{njt} includes alternative specific constants (ASCs) which allow for an intrinsic preference for each heating alternative – akin to a labelling effect[1]. In Equation (1) the taste parameters of utility functions are respondent specific. It is assumed that they follow distributions specified by a modeler such that $\beta_n \sim f\left(b + \Delta' z_n, \Gamma_n\right)$, with population mean b and variance-covariance matrix Γ_n. Moreover, it is possible to have the means of taste parameters to be influenced by observable respondent specific characteristics z_n and associated coefficient vector Δ.

[1] In general, not all features of a home heating system can be described based on chosen attributes (see Section 4), as there are numerous intangible heating system features (e.g., reputation, first-hand experience and space needs) which affect decisions. Since we do not observe these, they are captured only by the technology-specific ASCs.

The random utility model can be transformed into different classes of estimable choice models by making different assumptions about the error term. Since the utility function is ordinal, assumptions with respect to the error term variance may be expressed by scaling the utility function. To understand what scale means, note that the variance of extreme value one (EV1) type error in the MXL model is $\sigma^2\pi^2/6$, where the scale parameter σ is usually normalized to one to achieve identification.

A model that accounts for both preference and scale heterogeneity is the Generalized Mixed Logit (G-MXL) model (Fiebig et al. 2010; Greene and Hensher 2010). In the G-MXL model the random utility expression is

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt} = \left[\sigma_n \beta + \gamma\eta_n + (1-\gamma)\sigma_n \eta_n\right]' x_{njt} + \varepsilon_{njt}. \tag{2}$$

Here, β represents the population means and η represents individual specific deviations from these means. We also have the scale coefficient σ_n which is individual specific with $\sigma_n \sim LN(1\,,\,\tau)$, so that $\sigma_n = \exp(\bar{\sigma} + \tau\varepsilon_{0n})$ where $\varepsilon_{0n} \sim N(0\,,\,1)$. Note that $\bar{\sigma}$ denotes the population mean of scale and τ is the coefficient of the scale heterogeneity in the sample. For identification we need to normalize σ_n by setting $\bar{\sigma} = -\tau^2/2$.

In Equation (2), γ is a weighting parameter that indicates how the variance in residual preference heterogeneity varies with scale. If $\gamma = 1$, we get G-MXL-I model where $\beta_n = \sigma_n b + \eta_n$, whereas if $\gamma = 0$, we get G-MXL-II model where $\beta_n = \sigma_n(b + \eta_n)$. These are the two extreme cases of scaling residual taste heterogeneity.

Czajkowski et al. (2016) introduced how to account for both the systematic differences in the mean scale, and the systematic differences in its variance. In this paper we make the mean of the random scale parameter and its variance functions of respondent specific, perceived complexity-related covariates so that $\sigma_n \sim LN(1+\theta' h_n, \tau + \lambda' h_n)$. This further implies that the scale parameter is of the form

$$\sigma_n = \exp\left(\bar{\sigma} + \theta' h_n + \tau \exp(\lambda' h_n)\varepsilon_{0n}\right). \tag{3}$$

Above, h_n is a set of complexity-related covariates of individual *n* (that may overlap with z_n) and θ is the corresponding coefficient vector of covariates of mean scale, whereas λ is the corresponding coefficient vector of covariates of scale variance. As already noted, $\bar{\sigma}$ is a normalizing mean scale parameter.

Combining all terms, with *D* indicating draws from the predetermined distributions, the simulated log likelihood for the data is

$$\log L = \sum_{n=1}^{N} \log \frac{1}{D} \sum_{d=1}^{D} \prod_{t=1}^{T_n} P(j, X_{nt}, \beta_{ir})^{d_{njt}} , \tag{4}$$

where d_{njt} equals one if individual *n* makes chooses *j* in choice situation *t* and zero otherwise, and

$$P(j, X_{nt}, \beta_{nr}) = \exp\left(\beta'_{nr} x_{njt}\right) \Big/ \sum_{k=1}^{J} \exp\left(\beta'_{nr} x_{njt}\right). \tag{5}$$

In the Results section of this paper, we estimate the models discussed above for a stated choice experiment. As we want to focus on how the perceived choice complexity affects scale, we are particularly interested in the coefficients on h_n.

4. Case study

4.1 Survey design

The case study used in this paper is a Choice Experiment that investigates residential homeowner attitudes towards home heating systems in Finland. The final survey took place in August of 2014 and was executed via a mail questionnaire. Two thousand Finns were selected from the Population Information System of Finland. This sample was randomly drawn from a group of homeowners whose new detached houses had been finished between January of 2012 and May of 2014. Sampling was focused on individuals who had recently built new homes, since these individual were likely to be more familiar with alternative home heating systems, given that a CE based on the wide range of technologies available for domestic use might have imposed too great a cognitive burden on the general public. We received a total of 432 completed questionnaires implying a response rate of 21.6 percent.

For the final survey, we created 36 choice tasks and blocked these to six questionnaire versions. We used the Bayesian efficient D-optimal design in the Multinomial Logit framework (Ferrini and Scarpa 2007), where the prior parameter values were based on the priors obtained through the pilot survey. The heating system scenarios were designed to represent the most relevant primary and supplementary heating alternatives currently available in Finland[2]. The following six main heating alternatives were selected: district heating, solid wood fired boiler, wood pellet boiler, electric storage heating, ground heat

[2] The design of the survey instrument was started by identifying possible factors affecting individuals' heating mode choices based on previous literature (see e.g., Michelsen and Madlener 2012, 2013; Rouvinen and Matero 2013; Scarpa and Willis 2010). We also began discussions with experts (building authorities, civil engineers and researchers) to determine the most relevant main and supplementary heating technologies available today and the most important attributes with what we could describe these technologies in a realistic way.

pump (i.e., ground source heat pump) and exhaust air heat pump. Note that a labeled CE was the only way to represent realistic heating choice scenarios for the respondents, as each main heating alternative has label-specific attribute levels. The main heating systems were described using five attributes: supplementary heating systems, investment costs, operating costs, comfort of use and environmental friendliness. These are summarized in Table 1 and fully described in Ruokamo (2016).

// TABLE 1. (see the end of this paper)//

Six hypothetical choice tasks (see the example in Figure 1) were presented to each respondent. The respondents were asked to imagine that they were choosing a heating system for a new, 150 m^2 detached house, to compare the heating alternatives presented and then select the best alternative. Note that they were not asked to re-think the heating choice for their own, new home, since this would have been a difficult task for people.

// FIGURE 1. (see the end of this paper)//

4.2 Complexity-related covariates

As the aim of this paper is to explain differences in scale with alternative indicators of choice task complexity, we have to construct variables that measure different complexity-related aspects. These indicators of complexity were considered by respondents in the follow-up questions just after finishing the choice tasks.

First we focus on a variable that indicates how difficult the respondent perceived the choice tasks to be. The corresponding wording in the questionnaire was: “It was difficult to answer to the choice tasks presented to me.” Respondents then gave an answer indicating how much they agreed with this statement. The second variable is also closely related to perceived choice difficulty, but approaches it by linking the possibility of unrealistic choice alternatives with difficulties in answering. This statement was: "There were unrealistic choice alternatives that complicated answering." Based on these response items, we created two task complexity-related covariates: *Difficulty_General* and *Difficulty_Unrealistic*. Figure 2 presents distributions of these covariates. In both cases, respondents used a four-point Likert scale (1=strongly disagree, 2=somewhat disagree, 3=somewhat agree, 4=strongly agree). Also, a "do not know" option was included to allow respondents to state if they had no opinion or if they had not thought about a particular issue.

// FIGURE 2. (see the end of this paper)//

We strongly believe that these covariates are good proxies for perceived choice complexity. The covariates are only weakly correlated[3], and hence, they should measure different aspects of perceived choice complexity.

5. Results

The estimation was performed in Matlab, using 10 000 Pseudo random draws to simulate distributions of random parameters. We used multiple starting values to ensure convergence to a global maximum. We assigned normal distributions to random parameters, except for preferences towards operating and investment costs which were assumed to be log-normally distributed (so that people always prefer to pay less, ceteris paribus). A full list of determinants of the respondents' choices is presented in Table 2. ASC for district heat was normalized to zero in order to recognize relative preference rankings between main heating alternatives. The categorical attributes (supplementary

[3] The value of the correlation coefficients is the following:
Difficulty_General and *Difficulty_Unrealistic*: 0.3393

heating systems, comfort of use and environmental friendliness) were dummy coded in the analyses. Supplementary heating alternatives were compared with no supplementary heating as reference level. Comfort of use and environmental friendliness variables were compared with level good as reference level. Each complexity-related scale covariate was normalized so that its mean in the sample was 0 and standard deviation was 1. This enables us to examine differences between perceived complexities across individuals without encountering numerical problems in the estimation.

// TABLE 2. (see the end of this paper)//

We estimated the G-MXL model using 2508 observations. This model is reported in Table 3.

// TABLE 3. (see the end of this paper)//

We first focus on the preference parameters. Results show that all parameters are of the expected signs and are highly statistically significant. The statistical significance of the coefficient associated with the standard deviations of the random parameters indicates that they are significantly different from zero, and thus, the variables should be modelled as random. Differences in the mean coefficients of ASCs in the estimated model suggest that, on average, the respondents preferred ground heat and district heat systems over exhaust air heating pump, solid wood, wood pellet and electric storage heating systems with respect to other aspects not presented in the choice tasks. Results reveal that all three supplementary heating systems increase the choice probabilities of investigated main heating alternatives. Coefficients for operating costs and investment costs indicate that as operating and investment costs increase, the probability of choosing a system declines and utility levels decrease. The *Comfort_Excellent* coefficient measures change from the level "good" to "excellent". Therefore, when a coefficient has a positive sign, the described change increases the probability of selecting an alternative. Correspondingly, the *Comfort_Satisfactory* coefficient measures change from the level "good" to "satisfactory". When this coefficient is negative, the change decreases the probability of selecting an alternative. The *Environment_Excellent* and *Environment_Satisfactory* coefficients are interpreted in a similar way.

Next we focus on the analysis of scale[4]. In the G-MXL model, the τ coefficients representing the dispersion of individual scale coefficients is highly significant indicating significant heterogeneity in individual scale coefficients. The weighting parameter γ was constrained to 0 due to numerical problems in the estimation. In turn, we are using the G-MXL-II model in which the variance of residual taste heterogeneity is fully scaled.

Regarding the determinants of mean scale, results of the G-MXL show that the perceived difficulty of the choice tasks works as expected. Individuals who report that answering the choice tasks was difficult have less deterministic choices, i.e., mean scale decreases as perceived difficulty increases. Furthermore, if respondents report that unrealistic choice alternatives made answering more complicated, they seem to have lower mean scale.

We also allowed for the variance of individual scale to differ across respondents. The significant negative coefficient for *Difficulty_Unrealistic* indicates that respondents who found choice tasks unrealistic (and hence more complicated) have lower scale variance and thus are more similar to each other in terms of their randomness. However, the

[4] Note that while we do not report results here, we also estimated the Multinomial Logit (MNL), Scaled Multinomial Logit (S-MNL) and MXL models in which self-evaluated complexity variables were included as explanatory variables of scale. Comparing different approaches, the G-MXL model that allows for both preference and scale heterogeneity performs better than models that do not based on all information criteria (LL, McFadden R2, Ben-Akiva R2 and AIC). These results are available on the online annex.

explanatory power of scale variance turned out to be insignificant for *Difficulty_General* covariate.

Finally, we test if there are significant differences in preference parameters if we account for perceived difficulties. Table 4 reports results of the G-MXL models investigating preference and scale differences between respondents who found the choice tasks 'easy' and 'difficult'. We estimated three specifications where: in Model 1 means and variances of random parameters were assumed equal while scale could differ with respect to difficulty; in Model 2 means of random parameters were difficulty specific while holding variances constant; and in Model 3 both means and variances of random parameters were difficulty specific. Note that in Models 2 and 3 we divide respondents to two samples: the first group consists of individuals who reported that answering the choice tasks was easy or somewhat easy (208 individuals in total) and the other group consists of individuals who reported the opposite (207 individuals in total).

// TABLE 4. (see the end of this paper)//

To test the presence of statistically significant differences between 'easy' and 'difficult' sub-samples we present the Likelihood ratio test results in Table 5. By testing Model 1 vs. Model 2, Model 1 vs. Model 3 and Model 2 vs. Model 3 we cannot reject the equality hypothesis. Thus, preferences do not statistically differ between 'easy' and 'difficult' sub-samples in this dataset.

// TABLE 5. (see the end of this paper)//

6. Discussion and conclusion

The main result from this paper is that for two different measures of choice complexity in a home heating context, we find significant effects on scale heterogeneity. Individuals who stated that the choice tasks were "difficult" have less deterministic choices, so that scale decreases as perceived difficulty increases. Furthermore, if respondents reported that they found some choice tasks to be unrealistic, and therefore that choosing was more complicated, they seem to have lower mean scale as well as lower scale variance.

Even though the G-MXL model has been criticized by asking whether it is possible to identify separately both unobserved preference and scale heterogeneity (see Hess and Rose 2012), this modeling framework offers a way to address systematic shifts among individuals in the estimated parameters compared to the error term. The use of the G-MXL model requires, however, considerable effort to test alternative specifications and to ensure convergence to global maximum. For example, in our study about 10 000 draws were needed to simulate distributions of random parameters accurately and a state-of-art optimization method was used to find global optimal solutions.

Previous studies have found that the number of alternatives and attributes used to describe the good increase task complexity and scale heterogeneity (DeShazo and Fermo 2002; Hensher 2006; Louviere 2001; Swait and Adamowicz 2001a, 2001b). Despite of the fact that the choice tasks in this study were quite complicated (with six labeled heating systems and five attributes to describe the features of these heating systems), the results of the CE are robust involving statistically significant coefficients of attributes with expected signs (see also Ruokamo 2016).

In contrast to previous studies, we investigated scale heterogeneity by linking the differences in perceived choice complexity to mean scale and scale variance. The self-evaluated complexity-related covariates *Difficulty_General* and *Difficulty_Unrealistic* both have separate roles in explaining scale heterogeneity, even though they are correlated to some extent. Perceived choice complexity seems to be a multidimensional phenomenon. *Difficulty_General* captured the complexity of our choice tasks at general level, for example in terms of alternatives and attributes and their levels. On the other hand, some respondents considered the given CE incredible because some attribute levels seemed to be contradictory or unrealistic. This heterogeneity was captured by our *Difficulty_Unrealistic* covariate. Further, the specific description regarding *Difficulty_Unrealistic* covariate resulted in less variation when predicting respondents' preferences, i.e., in lower scale variance across individuals (see also Czajkowski et al. 2016).

Regarding welfare analysis, our results indicate that explicitly accounting for perceived choice complexity does not seem to affect preference parameters to a great degree in this data set. This indicates that, at least in our dataset, the bias resulting from failing to account for choice complexity may be small for welfare estimates. This may be due to the fact that the main heating systems and the attributes were carefully introduced to the respondents by including specific questions to ensure that respondents were familiar with the information provided before entering the choice tasks in the questionnaire. In addition, the target population in this study is expected to be very familiar with respect to the choice alternatives, as they had recently build new detached houses and made heating system choices in practice. Experience and familiarity have been identified being an important factor affecting scale heterogeneity also in previous studies (Czajkowski et al. 2014b; LaRiviere et al. 2014). In earlier studies controlling for different sources of scale heterogeneity on welfare estimates has been shown to vary widely. While Greene and Hensher (2010) and Czajkowski et al. (2014b) presented similar findings to ours, Kragt (2013) as well as Börger (2015) showed that failure to account for scale heterogeneity may induce significant biases in the estimated WTP confidence intervals.

Choice complexity is an important factor to be considered in designing and analyzing in choice experiments, in particular when respondents are not that familiar with the good at hand (Christie and Gibbons 2011; Fiebig et al. 2010). When we want to value complex goods such as domestic heating systems, we need to consider whether respondents participating in such studies are capable of revealing their true preferences. Respondents typically make more mistakes as the choice task becomes more complicated. Scale heterogeneity can be likely reduced by carefully testing the selected attributes and their levels as well as their descriptions in the questionnaire, but its presence cannot be totally avoided (see also DeShazo and Fermo 2002). However, the G-MXL model can be used to take into account the scale heterogeneity in the estimation. Further research is, nonetheless, needed to find the best practices to keep choice complexity within a minimum, and to handle uncertain responses in choice experiments.

References

Beck, M. J., Rose, J. M. and Hensher, D. A. (2013). Consistently Inconsistent: The Role of Certainty, Acceptability and Scale in Choice. *Transportation Research Part E* 56, 81–93.

Ben-Akiva, M., McFadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramaswamy, V., Rao, V., Revelt, D. and Steinberg, D. (1997). Modeling Methods for Discrete Choice Analysis. *Marketing Letters* 8(3), 272–286.

Boxall, P., Adamowicz, W. L. and Moon, A. (2009). Complexity in Choice Experiments: Choice of the Status Quo Alternative and Implications for Welfare Measurement. *The Australian Journal of Agricultural and Resource Economics* 53, 503–519.

Bush, G., Colombo, S. and Hanley, N. (2009). Should All Choices Count? Using the Cut-Offs Approach to Edit Responses in a Choice Experiment. *Environmental and Resource Economics* 44, 397–414.

Börger, T. (2015). Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment. Forthcoming in *Environmental and Resource Economics*.

Campbell, D., Hutchinson, W.G. and Scarpa, R. (2008). Incorporating dis-continuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics* 41, 401–417.

Czajkowski, M., Giergczny, M. and Greene, W. (2014a); Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. *Land Economics* 90(2), 324–351.

Czajkowski, M., Hanley, N. and LaRiviere, J. (2014b). The Effects of Experience on Preferences: Theory and Empirics for Environmental Public Goods. *American Journal of Agricultural Economics* 97(1), 333–351.

Czajkowski, M., Hanley, N. and LaRiviere, J. (2016). Controlling for the Effects of Information in a Public Goods Discrete Choice Model. *Environmental and Resource Economics* 63(3), 523–544.

Czajkowski, M., Kądziela, T. and Hanley, N. (2014c). We Want to Sort! – Assessing Households' Preferences for Sorting Waste. *Resource and Energy Economics* 36(1), 290–306.

Christie, M. and Gibbons J. (2011). The Effect of Individual ‘ability to choose‘ (scale heterogeneity) on the Valuation of Environmental Goods. *Ecological Economics* 70, 2250–2257.

Decker, T. and Menrad, K. (2015). House owners' Perception and Factors Influencing Their Choice of Specific Heating Systems in Germany. *Energy Policy* 85, 150–161.

DeShazo, J. R. and Fermo, G. (2002). Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. *Journal of Environmental Economics and Management* 44(1), 123–143.

Erdem, S., Campbell, D. and Thompson, C. (2014). Elimination and Selection by Aspects in Health Choice Experiments: Prioritising Health Service Innovations. *Journal of Health Economics* 38, 10–22.

Ferrini, S. and Scarpa, R. (2007). Designs with a Priori Information for Nonmarket Valuation with Choice Experiments: A Monte Carlo Study. *Journal of Environmental Economics and Management* 53, 342–363.

Fiebig, D. G., Keane, M. P., Louviere, J. and Wasi, N. (2010). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science* 29(3), 393–421.

Greene, W. H. and Hensher, D. A. (2010). Does Scale Heterogeneity across Individuals Matter? An Empirical Assessment of Alternative Logit Models. *Transportation* 37, 413–428.

Heiner, R. A. (1983). The Origin of Predictable Behavior. *The American Economic Review* 73, 560–595.

Hensher, D. A. (2006). How Do Respondents Process Stated Choice Experiments? Attribute Consideration under Varying Information Load. *Journal of Applied Econometrics* 21, 861–878.

Hess, S. and Rose, J. (2012). Can Scale and Coefficient Heterogeneity Be Separated in Random Coefficient Models? *Transportation* 39, 1225–1239.

Hess, S. and Stathopoulos, A. (2013). Linking Response Quality to Survey Engagement: A Combined Random Scale and Latent Variable Approach. *Journal of Choice Modelling* 7, 1–12.

Juutinen, A., Svento, R., Mitani, Y., Mäntymaa, E., Shojie, Y. and Siikamäki, P. (2012). Modeling Observed and Unobserved Heterogeneity in Choice Experiments. *Environmental Economics* 3(2), 57–65.

Keane, M. and Wasi, N. (2013). Comparing Alternative Models of Heterogeneity Consumer Choice Behavior. *Journal of Applied Econometrics* 28, 1018–1045.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy* 74, 132–157.

LaRiviere, J., Czajkowski, M., Hanley, N., Aanesen, M., Falk-Petersen, J. and Tinch, D. (2014). The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good. *Journal of Environmental Economics and Management* 68(2), 376–389.

Louviere, J. J. (2001). What if Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon. *Journal of Consumer Research* 28, 506–511.

Lundhede, T. H., Olsen, S. B., Jacobsen, J. B. and Thorsen, B. J. (2009). Handling Respondent Uncertainty in Choice Experiments: Evaluating Recoding Approaches Against Explicit Modelling of Uncertainty. *Journal of Choice Modelling* 2, 118–147.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. (Ed.). Frontiers in Econometrics. New York: Academic Press, 105–142.

Michelsen, C. C. and Madlener, R. (2012). Homeowners' Preferences for Adopting Innovative Residential Heating Systems: A Discrete Choice Analysis for Germany. *Energy Economics* 34, 1274–1283.

Michelsen, C. C. and Madlener, R. (2013). Motivational Factors Influencing the Homeowners' Decisions between Residential Heating Systems: An Empirical Analysis for Germany. *Energy Policy* 57, 221–233.

Regier, D. A., Watson, V., Burnett, H. and Ungar, W.J. (2014). Task Complexity and Response Certainty in Discrete Choice Experiments: An Application to Drug Treatments for Juvenile Idiopathic Arthritis. *Journal of Behavioral and Experimental Economics* 50, 40–49.

Revelt, D. and Train, K. (1998). Mixed Logit with Repeated Choices. *Review of Economics and Statistics* 80, 647–657.

Rouvinen, S. and Matero, J. (2013). Stated Preferences of Finnish Private Homeowners for Residential Heating Systems: A Discrete Choice Experiment. *Biomass and Bio-energy* 57, 22–32.

Ruokamo, E. (2016). Household Preferences of Hybrid Home Heating Systems – A Choice Experiment Application. *Energy Policy* 95, 224–237.

Scarpa, R., Gilbride, T. J., Campbell, D. and Hensher, D. A. (2009). Modelling Attribute Non-Attendance in Choice Experiments for Rural Landscape Valuation. *European Review of Agricultural Economics*, 36, 151–174.

Scarpa, R. and Willis, K. (2010). Willingness to Pay for Renewable Energy: Primary and Discretionary Choice of British Households' for Micro-generation Technologies. *Energy Economics* 32, 129–136.

Swait, J. and Adamowicz, W. (2001a). 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(2), 141–167.

Swait, J. and Adamowicz, W. (2001b). The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research* 28, 135–148.

Thurstone, L. L. (1927). A Law of Comparative Judgement. *Psychological Review* 4, 273–286.

Train, K. (2009). Discrete Choice Methods with Simulation (2nd Ed.). Cambridge University Press, Cambridge.

Table 1. Attributes and levels.

ATTRIBUTE	DESCRIPTION	LEVELS
Supplementary heating system	Supplementary heating system works alongside the main heating system.	*District heat*: no supplementary heating systems *Others*: Level 1: no supplementary heating systems Level 2: solar panel and solar water heater Level 3: water fireplace Level 4: outside air heat pump
Investment cost (€)	The investment cost includes costs associated with the heating device and installation as well as space requirements.	*District heat*: 6000€, 7500€, 9000€, 10500€ *Solid wood fired*: 4500€, 7000€, 9500€, 12000€ *Wood pellet*: 8000€, 11000€, 14000€, 17000€ *Electric storage heating*: 6000€, 8500€, 11000€, 13500€ *Ground heat pump*: 13000€, 16000€, 19000€, 22000€ *Exhaust air heating pump*: 7000€, 9000€, 11000€, 13000€
Operating cost (€/year)	The operating cost includes heating system's annual electricity/fuel consumption and maintenance costs.	*District heat*: 800€, 1000€, 1200€, 1400€ *Solid wood fired*: 600€, 850€, 1100€, 1350€ *Wood pellet*: 750€, 950€, 1150€, 1350€ *Electric storage heating*: 1050€, 1350€, 1650€, 1950€ *Ground heat pump*: 500€, 650€, 800€, 950€ *Exhaust air heating pump*: 800€, 1000€, 1200€, 1400€
Comfort of use	The comfort of use describes the required work to ensure the faultless operation of the heating system, e.g., cleaning and adjusting the device and adding fuel.	*Solid wood fired and wood pellet*: satisfactory, good *District heat, electric storage heating, ground heat pump and exhaust air heating pump*: good, excellent
Environmental friendliness	The environmental friendliness describes the ecological facts associated with each available heating system.	*District heat, solid wood fired, wood pellet and ground heat pump*: good, excellent *Electric storage heating and exhaust air heating pump*: satisfactory, good

Table 2. Definition of explanatory variables.

Variable	***Type***
Preference parameters	
ASC_Ground heat pump	Dummy-coded
ASC_Exhaust air heat pump	Dummy-coded
ASC_Solid wood	Dummy-coded
ASC_Pellet wood	Dummy-coded
ASC_Electric storage	Dummy-coded
Supplementary_Solar	Dummy-coded
Supplementary_Water-fireplace	Dummy-coded
Supplementary_Outside air heat pump	Dummy-coded
Investment cost (-/10000 €)	Continuous
Operating cost (-/1000 €/year)	Continuous
Comfort_Satisfactory	Dummy-coded
Comfort_Excellent	Dummy-coded
Environment_Satisfactory	Dummy-coded
Environment_Excellent	Dummy-coded
Covariates of scale	
Difficulty_general	Continuous
Difficulty_unrealistic	Continuous

Table 3. Results of the G-MXL model investigating unobserved and observed scale heterogeneity with respect to perceived choice task complexity.

	Distribution	**Mean (s.e.)**	**Standard deviations (s.e.)**
Preference parameters			
ASC_Ground heat pump	Normal	2.2742*** (0.4764)	3.8064*** (0.5820)
ASC_Exhaust air heat pump	Normal	-1.7802*** (0.4227)	3.2861*** (0.5880)
ASC_Solid wood	Normal	-5.4257*** (0.7843)	6.2769*** (0.9578)
ASC_Pellet wood	Normal	-3.3976*** (0.5791)	0.8984 (0.5974)
ASC_Electric storage	Normal	-3.6477*** (0.7237)	4.2947*** (0.8081)
Supplementary_Solar	Normal	1.2313*** (0.2785)	1.9969*** (0.3480)
Supplementary_Water-fireplace	Normal	0.4406* (0.2453)	2.0002*** (0.3978)
Supplementary_Outside air pump	Normal	0.7000*** (0.2480)	1.1197*** (0.2805)
Investment cost (-10000 EUR)	Log-normal[5]	1.5103*** (0.1536)	0.7103*** (0.0643)
Operating cost (-1000 EUR/y)	Log-normal	2.0174*** (0.1444)	0.6043*** (0.0467)
Comfort_Satisfactory	Normal	-5.2395*** (1.0384)	4.5725*** (0.7884)
Comfort_Excellent	Normal	0.5019*** (0.1534)	0.7599*** (0.2558)
Environment_Satisfactory	Normal	-2.6409*** (0.5696)	1.9341** (0.8475)
Environment_Excellent	Normal	0.9259*** (0.2002)	1.2019*** (0.2650)
Scale parameters			
τ (G-MXL scale variance)	Log-normal	2.3801*** (0.7306)	
Difficulty_General		-0.1633* (0.0906)	0.0898 (0.0613)
Difficulty_Unrealistic		-0.2092* (0.1100)	-0.2688** (0.1059)
Model diagnostics			
LL at constant(s) only	-3754.94		
LL at convergence	-2779.93		
McFadden's pseudo-R^2	0.259661		
Ben-Akiva-Lerman's pseudo-R^2	0.373578		
AIC/*n*	2.243526		
n (observations)	2508		
r (respondents)	418		
k (parameters)	33		

*, **, *** indicate significance at 0.1, 0.05, 0.01 level, respectively.

[5] For log-normal distributions the parameters of the underlying normal distribution are presented.

Table 4. Results of the G-MXL model investigating preference and scale differences between respondents who found the survey 'easy' and 'difficult'.

	Model 1 **G-MXL with preference parameters equal for both sub-samples**		**Model 2** **G-MXL with preference parameter means interacted with a 'difficult' dummy**			**Model 3** **G-MXL with sub-sample specific preference parameters**			
						Answering 'easy'		Answering 'difficult'	
	Mean (s.e.)	S.D. (s.e.)	Mean – main effect (s.e.)	Mean – 'difficult' sub-sample shifter (s.e.)	S.D. (s.e.)	Mean (s.e.)	S.D. (s.e.)	Mean (s.e.)	S.D. (s.e.)
Preference parameters									
ASC_Ground heat pump	2.9352*** (0.7798)	5.0296*** (0.9937)	1.9087*** (0.4336)	-0.1034 (0.5757)	3.1038*** (0.3914)	2.6114*** (0.6711)	3.9848*** (0.7165)	1.5517*** (0.4066)	2.7578*** (0.4153)
ASC_Exhaust air heat pump	-2.2345*** (0.6643)	3.8434*** (0.8797)	-1.2573*** (0.4516)	-0.3035 (0.5844)	2.6037*** (0.3925)	-1.4659*** (0.5613)	3.3598*** (0.7539)	-1.4768*** (0.4173)	2.1259*** (0.4630)
ASC_Solid wood	-7.5416*** (1.7489)	7.7354*** (1.6475)	-5.5308*** (0.8069)	1.0285 (0.7230)	5.7956*** (0.8005)	-7.5601*** (1.5369)	6.4340*** (1.2012)	-4.4414*** (0.8772)	4.9982*** (0.8722)
ASC_Pellet wood	-4.5291*** (1.0356)	0.8076 (0.6698)	-3.8223*** (0.6732)	0.8161 (0.6365)	1.8410*** (0.6207)	-4.3500*** (0.9791)	2.0233** (0.8060)	-2.1583*** (0.4410)	0.8592 (0.7531)
ASC_Electric storage	-3.8509*** (0.9888)	4.6309*** (1.1625)	-1.8524*** (0.5374)	-1.5589** (0.6729)	2.9378*** (0.4806)	-2.4926*** (0.7710)	4.3348*** (0.9332)	-3.3631*** (0.7152)	2.7787*** (0.5403)
Supplementary_ Solar	1.5403*** (0.4134)	2.6588*** (0.6015)	1.2976*** (0.3051)	-0.5793 (0.4145)	1.7639*** (0.2785)	1.7235*** (0.4424)	1.5330*** (0.3708)	0.5566** (0.2806)	1.8301*** (0.3416)
Supplementary_ Water-fireplace	0.5141 (0.3297)	2.3243*** (0.5747)	0.4834* (0.2934)	-0.2646 (0.4131)	1.5737*** (0.3016)	0.5517 (0.3529)	2.0928*** (0.5461)	0.2101 (0.2789)	1.3163*** (0.4275)
Supplementary_ Outside air pump	0.8744** (0.3611)	1.5223** (0.6705)	0.4283 (0.3027)	0.1651 (0.4025)	0.9186** (0.3900)	0.4878 (0.3503)	1.4610*** (0.4450)	0.6310** (0.2488)	0.5412 (0.4931)
Investment cost (-10000 EUR)	1.7371*** (0.2149)	0.6903*** (0.0710)	1.9191*** (0.1189)	-0.1709 (0.1248)	0.6381*** (0.0638)	1.5428*** (0.1968)	0.7591*** (0.0902)	1.1807*** (0.1449)	0.6709*** (0.0817)
Operating cost (-1000 EUR/y)	2.2566*** (0.1992)	0.5939*** (0.0637)	1.3185*** (0.1304)	-0.0551 (0.1336)	0.6944*** (0.0836)	2.2071*** (0.1695)	0.4891*** (0.0820)	1.5820*** (0.1484)	0.7834*** (0.0924)
Comfort_Satisfacto ry	-6.8094*** (1.7478)	6.2441*** (1.4491)	-3.6709*** (0.7297)	-0.5766 (0.8269)	3.7346*** (0.6197)	-5.8842*** (1.4195)	6.0564*** (1.2354)	-2.6773*** (0.5622)	0.9337 (1.0220)
Comfort_Excellent	0.5410*** (0.2013)	0.6652 (0.5391)	0.1844 (0.1765)	0.4312* (0.2483)	0.7729*** (0.2502)	0.2592 (0.1989)	0.6963 (0.4679)	0.4553*** (0.1740)	1.0109*** (0.2854)
Environment_Satisf actory	-3.3151*** (0.8367)	2.8746*** (0.9097)	-2.3201*** (0.5327)	-0.0729 (0.6203)	2.1945*** (0.5090)	-2.7937*** (0.7901)	2.0525*** (0.7773)	-2.1136*** (0.5970)	1.9890*** (0.5421)
Environment_Excel lent	1.0317*** (0.2783)	1.7685*** (0.4522)	0.6982*** (0.1955)	0.0752 (0.2728)	1.0800*** (0.2219)	0.8700*** (0.2607)	1.3000*** (0.3771)	0.6477*** (0.1895)	1.0164*** (0.2783)
Scale parameters									
τ	0.8181 (0.7819)		3.6823*** (0.5348)	-	-	3.9478*** (0.6489)	-	-	-
Difficulty_General	-0.4825** (0.1951)	1.0023 (0.6353)	-	-	-	-	-	-	-
Model diagnostics									
LL at constant(s) only	-3708.35		-3708.35			-3708.35			
LL at convergence	-2737.59		-2730.90			-2721.10			
McFadden's pseudo-R²	0.2618		0.2636			0.2662			
Ben-Akiva-Lerman's pseudo-R²	0.3757		0.3761			0.3791			
AIC/*n*	2.2349		2.2395			2.2433			
n (observations)	2478		2478			2478			
r (respondents)	415		415			415			
k (parameters)	31		43			57			

*, **, *** indicate significance at 0.1, 0.05, 0.01 level, respectively.

Table 5. Likelihood ratio tests results investigating the presence of statistically significant differences between 'easy' and 'difficult' sub-samples.

	Test statistic	Degrees of freedom	P-value
Model 1 vs. Model 2 (differences in means)	13.3844	12	0.3417
Model 1 vs. Model 3 (differences in means and variances)	32.98942	26	0.1624
Model 2 vs. Model 3 (differences in variances)	19.60502	14	0.1431

Figure 1. Example of a choice task.

As a reminder: the heating system is chosen for new 150 m^2 sized detached house						
CHOICE TASK 1	***Ground heat***	***Exhaust air heating pump***	***Solid wood fired***	***Wood pellet***	***Electric storage heating***	***District heat***
Supplementary heating system	Solar panel and solar water heater	Water - circulating fireplace	No supplementary heating systems	Outside air heat pump	Water-circulating fireplace	No supplementary heating systems
Investment cost (€)	16000	7000	7000	17000	8500	9000
Operating cost (€/year)	650	1400	1100	1350	1350	800
Comfort of use	Good 😐	Excellent ☺	Satisfactory ☹	Satisfactory ☹	Good 😐	Excellent ☺
Environmental friendliness	Excellent ☺	Satisfactory ☹	Excellent ☺	Good 😐	Good 😐	Good 😐
I CHOOSE:	☐	☐	☐	☐	☐	☐
Choose the best alternative by ticking one of the above boxes.						

Figure 2. Distributions of task complexity related covariates.