Journal of Agriculture and Food Research 29 (2026) 102998

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Comparing Generalized Multinomial Logit Model and Random Parameter Logit Model in Preference Space and in Willingness-to-Pay Space

Xianwen Chen *

Department of Business Administration, Inland School of Business and Social Sciences, University of Inland Norway, Norway

A R T I C L E I N F O

JEL classification:
C52
D12
Q22

Keywords:
Mixed multinomial logit model
Generalized multinomial logit model
Random parameter logit model
Willingness to pay space
Choice experiment
Ecolabel
Fish

A B S T R A C T

The Generalized Multinomial Logit (GMNL) model and the Random Parameter Logit (RPL) model are two specifications of the Mixed Multinomial Logit (MMNL) model. These models can be estimated in preference space or in Willingness-to-Pay (WTP) space. We use data from a French stated choice experiment concerning the WTP for ecolabeled wild fish and organic-labeled farmed fish. The data is estimated using GMNL and RPL models in preference space and in WTP space. We find that our participants were willing to pay around 10% premiums for ecolabel and for organic label in France in 2008. Theoretically, estimation in WTP space has the advantages of assuring finite moments and avoiding the potential problem of fat tail in WTP distribution. We find mixed evidence from the Kernel Density Estimation (KDE) of WTP, such that our results do not conclusively support these theoretical advantages of estimation in WTP space. We also compare model fit and WTP estimates across the four models. The GMNL model in preference space has the best fit, but the differences in model fit among the four models are small. With one exception, the estimated 95% confidence intervals of WTP from the four models overlap. The similarity of the results from the four models demonstrates robustness of the estimates across model specifications. The similarities in model fit and in WTP estimates indicate that all four models have worked equally well in our dataset. We discuss model fit and WTP estimation and provide suggestions on choosing between the GMNL model and the RPL model, from the perspectives of model fit, model parsimony, economic meaningfulness of differences in WTP estimates, and practicality.

1. Introduction

We make decisions all the time on virtually everythingon topics related to agriculture and food, and beyond. We make decisions such as which groceries to buy, what food to have for dinner, what kind of property to purchase, what animals to farm on the ranch, and what crops to grow on the land. We also make decisions outside of agriculture, such as where to live, what education to have, which person to date, which career to pursue, which stock to invest in, and which president candidate to vote for. Logit and probit models are useful mathematical approximations to model our decision making, backed by the theoretical foundation of McFadden's Random Utility Maximization (RUM). The empirical estimation of choice data is usually referred to as Discrete Choice Modeling (DCM) today. Recent advances in this field include, for example, random parameterization of coefficients, latent class parameterization of coefficients, accounting for scale heterogeneity in the error term, estimation in Willingness-to-Pay (WTP) space, regret minimization, and inattention.

This paper compares the Random Parameter Logit (RPL) model with the Generalized Multinomial Logit (GMNL) model [1,2] in preference space and in Willingness-to-Pay (WTP) space. When analyzing data, we constantly face an increasing number of empirical models that we have to choose from. It is therefore of interest and importance to compare models, to shed light on whether a model greatly outperform another model, or whether two models are comparable that they produce comparable results.

The Mixed Multinomial Logit (MMNL) model has replaced the Multinomial Logit (MNL) model and become the standard model for estimating choice data [3–6]. While in theory one may use complex parameterizations and residual distributions in a MMNL model, researchers commonly use normal, lognormal, or truncated normal distributions for most or all of the parameters. We refer to such specifications of the MMNL model as the RPL model.

The key advantage of the RPL model is that it captures taste heterogeneity. An MNL model estimates the parameter for an attribute. The estimated parameter represents the mean weight of the attribute,

* P.O. Box 400, 2418, Elverum, Norway.
E-mail addresses: xianwen.chen@inn.no, xchen@live.no.

https://doi.org/10.1016/j.jafr.2026.102998
Received 16 December 2025; Received in revised form 5 May 2026; Accepted 13 May 2026
Available online 2 June 2026

measuring the mean preference for the attribute, applicable for all respondents in the dataset. Consequently, we cannot measure individual differences in the preference for this attribute across respondents using an MNL model, which is a clear disadvantage because people usually, if not universally, prefer the same attribute differently. In contrast, the RPL model can estimate the distribution of the parameter for an attribute. The mean of the estimated distribution is the mean utility weight for the attribute, which is conceptually equivalent to the estimated parameter of an MNL model. However, because the parameter is now a distribution in the RPL model, the estimated distribution captures how the utility weight for the attribute varies across the respondents of the study. We call such variation in utility weight for the attribute, or we may say variation in preference for the attribute, the taste heterogeneity for the attribute among respondents. By capturing the variations in taste (preference) across individuals, the RPL model not only drastically increases model fit as compared to the MNL model, but also captures a critically important aspect of human nature, which is the differences in the preference for an attribute across individual persons.

However, a standard RPL model cannot identify another important source of individual differences, the scale heterogeneity. Scale heterogeneity refers to systematic differences across individuals in the overall variability of the unobserved component of utility, holding tastes for attributes constant. In a logit framework, it shows up as differences in the scale of the error term. Some respondents make very sharp and deterministic choices. Their choices are associated with high scale and low error variance. However, other respondents' choices may be more random or inconsistent. Their choices are associated with low scale and high error variance, even if their underlying preferences over attributes are similar to the respondents that choose more deterministically.

Accounting for scale heterogeneity is important. If scale heterogeneity exists in the dataset but is absent in the model, part of what is actually variation in decision consistency can be misinterpreted as taste heterogeneity, leading to distorted inferences about how much preferences truly differ across individuals. In applications where forecasting or policy evaluation relies on accurately capturing how sharply people respond to changes in attributes or prices, omitting scale heterogeneity will likely degrade both predictive performance and the credibility of WTP and welfare analyses.

The GMNL model was developed with the hope to allow for separate estimation of taste heterogeneity and scale heterogeneity in choice data [1,2], thus providing a model framework that captures both types of heterogeneities. However, later research suggests that the GMNL model was not able to disentangle the two types of heterogeneities from each other [7] and that GMNL is a special type of the MMNL model [8].

Nevertheless, the GMNL model often outperforms the RPL model in terms of model fit. We have seen evidences from studies in marketing (e.g., Ref. [1]), transportation (e.g., Ref. [2,7,9,10]), resource and environmental economics (e.g., Ref. [11–16]), health economics (e.g., Ref. [17–19]), agricultural economics (e.g., Ref. [20–27]), labor economics (e.g. Ref. [28,29]), and voting behavior research (e.g. Ref. [30]).

Traditionally, logit models, including the MMNL model, are usually estimated in preference space, where estimated distribution of a parameter represents the distribution of the utility weight of the attribute (see e.g. Ref. [31]). It is called estimation in preference space, because the estimated parameters are in the unit of utility, util. For example, it has been a common practice to estimate choice data using the RPL model in preference space in applied studies (e.g., Ref. [6, 32–34]).

A potential issue with estimating an MMNL model in preference space is that an estimated WTP distribution may have a large variance and may not have finite moments [35]. The mechanics are the following. In a model in preference space, distributions of the parameters are specified in utility units. When estimating a WTP distribution for a non-price attribute, the distribution of the parameter for the non-price attribute will be divided by the distribution of the price parameter. A potential issue is that the ratio of two distributions may have large variance and may not have finite moments.

Alternative to estimating in preference space, estimating the MMNL model in WTP space theoretically does not have such problems [36,37]. Train and Weeks [36] and Sonnier, Ainslie and Otter [37] showed that by setting the utility weight of the price parameter to the same as the scale parameter, the utility function in the RPL model is rescaled. As a result, the parameter estimate of each product attribute is in a monetary unit (e.g., US dollar) and equals the marginal WTP for the attribute. Hence, the WTP values are estimated directly [36]. Therefore, this parameterization is referred to as estimating in WTP space.[1] When developing the GMNL model, Fiebig et al. [1] showed that under a specific set of parameter restrictions, the GMNL model can also be parameterized and estimated in WTP space.

Another advantage that estimation in WTP space has over estimation in preference space is the simplification of the estimation procedure of WTP. When estimating a WTP distribution in preference space, we need to use methods such as the Delta method or the Krinsky–Robb method, to obtain the WTP estimates and their associated standard errors [38]. This is no longer necessary when estimating in WTP space. This simplification represents an advantage for the applied researchers.

However, estimation in WTP space also has disadvantages. Rescaling the price parameter results in a utility model that is nonlinear in the random parameters. Train and Weeks [36] and Hole and Kolstad [39] found that estimation of the RPL model in WTP space gave a small reductions in the Log-Likelihood Function Value (LLFV) as compared with estimation in preference space. However, Scarpa, Thiene and Train [33] reached the opposite conclusion.

Our primary objective of this study is to compare results from the GMNL and the RPL models estimated in preference space and in WTP space. We compare model fit and WTP estimates between the GMNL model and the RPL model and between estimation in preference space and in WTP space. The insights on model robustness and model comparison that we derive in the primary objective are not time sensitive. The paper fits timely in the current methodological discussion in the literature on the empirical performance of different MMNL model specifications.

Our secondary objective is to investigate French consumers' WTP for several wild and farmed fish and the premiums for fish ecolabels in France, by estimating models using data from a stated choice experiment. France is an interesting case because it is one of Europe's largest seafood markets. The WTP estimates, however, reflect the conditions of 2008. They have historical values but may not reflect the current market conditions or current consumer preferences.

2. Empirical background

A number of ecolabeling programs have been introduced following increased consumer concerns about overexploitation and seafood production issues such as safety, quality, environmental effects, sustainability, and animal welfare [40–50]. The long-term success of these ecolabeling schemes depends on firms' compliance and consumers' acceptance of them. The most important success measure whether, and how much, consumers are willing to pay in price premiums for the labeled products.

The certification program of the Marine Stewardship Council (MSC) is currently the most widely used and recognized sustainable wild fish labeling scheme in the world [51,52], which is also used in France. As of September 2025, 592 fisheries had already been certified by the MSC program, covering 18.9% of global marine catch, and MSC labeled seafood is sold by leading retailers all over the world [53].

However, it is difficult for consumers to know which fish stocks are depleted. It may not be enough to know what type of fish is being

[1] In health economic literature, estimation in WTP space is sometimes termed as estimation in Quality-Adjusted Life Year (QALY) space [69].

purchased. In many cases, the consumer needs to know where and when the fish was caught. Take Norwegian cod in early 2010s for example. The cod fishery in the Barents Sea was generating record landings and had been granted the MSC label, while at the same time cod from the North Sea and the Norwegian coast was considered to be under considerable pressure and had not been granted the MSC label. The perplexing detail was that the cod from the Barents comes to the shores of Northern Norway during the winter months to spawn. Hence, the environmentally concerned fish consumer must look for the MSC label to avoid buying wild fish from a depleted stock.

A few ecolabeling programs for farmed fish have been gaining international acceptance in recent years, although none of them have so far gained recognition at the level of the MSC label [54]. The Aquaculture Stewardship Council (ASC), founded in 2009 by the World Wildlife Fund and the Dutch Sustainable Trade Initiative, is the aquaculture version of the MSC. Another international ecolabel example is Global Aquaculture Alliance's Best Aquaculture Practices Certification. There are also national ecolabels in some countries. For example, the Agriculture Biologique (AB) label is the most widely used ecolabel for food in France, which certifies food products with an organic content of at least 95%. Farmed fish can be labeled as organic, whereas wild fish cannot. However, we were unable to find any certified organic fish products in the French market in 2008, at the time of the choice experiment.

From numerous studies of ecolabeling's effects on wild and farmed fish, labeling is generally found to have a positive effect. Jaffry et al. [42] used a choice experiment and found that ecolabeled seafood from a sustainably managed fishery has up to a 7% higher probability of being chosen by participants. Roheim, Asche, and Santos [55] analyzed scanner data of MSC-certified frozen processed Alaskan pollock products and found that UK consumers are willing to pay a 14% premium for the label. Olesen, Myhr, and Rosendal [56] conducted a non-hypothetical choice experiment and found that the average Norwegian consumer is willing to pay a 15% premium for organic salmon. Mauracher, Tempesta, and Vecchiato [57] found a significant price premium for organically bred Mediterranean sea bass. Uchida et al. [46] found that ecolabels have positive and significant effects on both wild and farmed fish. Bronnmann and Asche [48] found that ACS labeled farmed salmon was having a price premium that it was valued similar to MSC labeled wild salmon by German consumers.

3. Methods

In this section we describe the experimental design and econometric specifications used to address our two objectives: (i) comparing the empirical performance of GMNL and RPL models in preference and in WTP space and (ii) obtaining robust WTP estimates for ecolabeled and organic fish. The choice experiment was hypothetical and implemented as a standard labeled discrete choice design as commonly used in food and environmental economics, with attributes chosen to mirror French retail markets at the time. Real fish was presented to the participants in the experiment. On the econometric side, we focus on RPL and GMNL models, because RPL is the standard logit model specification for DCM and because GMNL has demonstrated better model fit than RPL, potentially due to its capture of scale heterogeneity. We estimate the models in preference and in WTP space, to further discern the performance of estimation in WTP space, as compared to the traditional method of estimation in preference space.

3.1. Experimental design

The experiment was carried out in the sensory laboratory of L'Institut National de la Recherche Agronomique (INRA) in Dijon in December 2008. Potential participants were randomly drawn from INRA's database of people who were willing to participate in food studies. They were asked to complete a short survey about their consumption and purchasing frequencies of fish products. We recruited participants that either ate fish at least once every month or bought fish at least once every two months. This means that we included people that consumed and/or bought fish at least rarely. We excluded people that almost never consumed and almost never bought fish. Each participant was paid €25.

The choice experiment was based on a choice design constructed by the SAS macro MktEx with zero priors. The D-efficiency of the total design was 96.52. The design used 112 choice sets organized into seven blocks. Each block therefore consisted of 16 choice sets. The order of choice sets was randomized within each block. Each block of choice sets was used in two sessions. Each participant was randomly assigned to one of the sessions.

The eight fish products used in the experiment are listed in Table 1. Salmon and cod are among the most frequently purchased fish species in France. Monkfish was included as an expensive substitute. Pangasius was included as an inexpensive substitute. Wild cod was either unlabeled or labeled as ecological. Farmed cod and farmed salmon were either unlabeled or labeled as organic. Monkfish and pangasius were not labeled. We labeled ecological fish with the MSC label and organic fish with the AB label.

To increase the realism of the experiment, the alternatives were presented in 300-g packages of real fish. Each choice set included three packages that were displayed in a Styrofoam box filled with ice. The fish varied with respect to the species, whether it was wild or farmed, its price, the use of the ecolabel, and its area of origin. A “none-of-these alternatives” was included as an option, for participants to opt out.

Following the experiment, each participant was asked to complete a questionnaire that included questions about attitudes toward fish in general as well as those fish species included in the experiment, demographics, and attitudes toward labels. The final data set consisted of 2300 choices made by 144 participants.

3.2. Econometric models

The GMNL and RPL models in preference and WTP spaces are presented and the specification and estimation of the models are described in the following subsections.

3.2.1. The GMNL and RPL models in preference space

We specify a linear utility function. Participant n chooses alternative j in choice situation t and obtains utility U_{njt}:

$$U_{njt} = \boldsymbol{\beta}_n \mathbf{X}_{njt} + \varepsilon_{njt}, \tag{1}$$

where $\mathbf{X}_{njt}$ includes the price p_{njt} and a vector of non-price attributes $\mathbf{x}_{njt}$, $\boldsymbol{\beta}_n$ is the random parameter vector for participant n, and ε_{njt} is an

Table 1
The products in the experiment.

Species	Wild or Farmed	Ecolabel	Area of Origin[a]	Price Range 300 Grams[b]
Salmon	Farmed	No	Norway	€1.95-5.45
Salmon	Farmed	Organic AB	Norway	€3.45-7.95
Cod	Farmed	No	Norway	€2.95-6.95
Cod	Farmed	Organic AB	Norway	€4.95-10.95
Cod	Wild	No	North Atlantic	€2.95-6.95
Cod	Wild	MSC	North Atlantic	€4.95-10.95
Monkfish	Wild	No	North Atlantic	€5.45-11.45
Pangasius	Farmed	No	Vietnam	€1.45-4.95

Notes.
[a] The origins of the different species are the origins that were most common in the French market. For the organic cod and salmon, we used the same origin as for the conventional cod and salmon.
[b] An eight-point price scale was used for each product.

idiosyncratic error term that is assumed to follow an extreme value distribution. The non-price attributes are specified as dummy variables. The dummy variables are used to specify the eight products in Table 1. There are five types of fish without ecolabels (farmed salmon, farmed cod, wild cod, monkfish, and farmed pangasius) and three types of fish with ecolabels (AB-labeled farmed salmon, AB-labeled farmed cod, and MSC-labeled wild cod).

In the GMNL model in preference space, the random parameter vector in equation (1) is specified as:

$$\boldsymbol{\beta}_n = \sigma_n \boldsymbol{\beta} + [\gamma + \sigma_n(1-\gamma)]\boldsymbol{\Gamma}\mathbf{w}_n, \text{where } \mathbf{w}_n \sim \mathrm{N}[\mathbf{0}, \mathbf{I}] \text{ and } 0 \le \gamma \le 1, \tag{2}$$

where.

- $\sigma_n = \exp\left(\tau \nu_n - \tau^2/2\right)$ represents scale heterogeneity and follows a standard lognormal distribution, where $\nu_n \sim \mathrm{N}[0,1]$ and $\tau \in [0,1]$,
- $\boldsymbol{\beta}$ is the vector that contains means of the distribution of $\boldsymbol{\beta}_n$,
- γ is the weighting parameter that $0 \le \gamma \le 1$,
- $\boldsymbol{\Gamma}$ is the lower triangular Cholesky matrix,
- and $\mathbf{w}_n$ is the vector of random variables with zero means and unit variances. The elements of $\mathbf{w}_n$ are uncorrelated.

Within the scale heterogeneity σ_n, τ governs the extent of scale heterogeneity. If there is no scale heterogeneity, the estimated τ, $\hat{\tau}$, will be zero. If $\hat{\tau}$ is zero, σ_n becomes one for all participants, implying that there is no scale heterogeneity across participants. If $\hat{\tau}$ is not zero, scale heterogeneity is found. The larger $\hat{\tau}$ is, the larger the scale heterogeneity is.

In the general GMNL specification in (2), the weighing parameter γ weighs how the variance of residual taste heterogeneity varies with scale heterogeneity when both types of heterogeneities are included [1]. When $\gamma = 0$, the model allocates scale heterogeneity separately from the random taste coefficients. When $\gamma = 1$, scale heterogeneity multiplies both the mean and the deviations of the random coefficients. Intermediate values of γ between zero and one represent mixtures of these two corner cases. Thus, γ governs whether scale heterogeneity is modeled as affecting only the overall error variance, or also amplifying or dampening individual taste deviations around the mean. Keane and Wasi [58] argued that γ does not have to be restricted between zero and one. Their argument was adopted by Gu, Hole, and Knox [59] when they implemented the GMNL estimation program for Stata. We, however, follow Fiebig et al. [1], Greene and Hensher [2], and Greene [38] and restrict $\gamma \in [0,1]$. Notice that this difference only affects the estimation of the GMNL model in preference space, because the GMNL model in WTP space restricts γ to be zero.

The GMNL model in equation (2) collapses into the RPL model in preference space when τ is zero (thus $\sigma_n = 1$). The random parameter vector in equation (2) becomes:

$$\boldsymbol{\beta}_n = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_n. \tag{3}$$

Let α_n be the parameter associated with the price. From equation (1), we derive the WTP distributions for the non-price attributes as the marginal rate of substitution between the non-price attributes and the price, or:

$$\mathbf{WTP}_n^{np} = \frac{\boldsymbol{\beta}_n^{np}}{-\alpha_n}, \tag{4}$$

where $\mathbf{WTP}_n^{np}$ is the vector of the nth participant's marginal WTP values for the seven non-price attributes and $\boldsymbol{\beta}_n^{np}$ is the vector of the seven non-price random parameters.

3.2.2. The GMNL and RPL models in WTP space

The RPL model using a money metric utility function was first estimated by Train and Weeks [36]. The estimated parameter for an attribute is the marginal WTP value for the attribute. Train and Weeks [36] described the model as a model estimated in WTP space.

Fiebig et al. [1] showed that the GMNL model also can be estimated in WTP space. By fixing the weighting parameter γ to be zero, equation (2) becomes:

$$\boldsymbol{\beta}_n = \sigma_n(\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{w}_n) \tag{5}$$

As compared with the RPL model in preference space given by equation (3), the model represented by equation (5), includes the multiplier of σ_n, which is the scale heterogeneity parameter.

By restricting $\gamma = 0$, the estimation in WTP space explicitly implies that the scale heterogeneity does not interact with residual taste heterogeneity. As a consequence, should there be interaction between scale heterogeneity and residual taste heterogeneity, GMNL in WTP space cannot capture it. Together with the additional nonlinearity introduced by normalizing the price coefficient to the scale parameter, this makes the GMNL model in WTP space less flexible in capturing certain patterns in the choice data, should such patterns exist.

Restricting $\gamma = 0$ is necessary because we need to normalize the parameter for price, so that the estimated parameters for other variables will be interpreted directly as marginal WTP values. If $\gamma \neq 0$ in WTP space, we would create additional nonlinear interactions between scale heterogeneity and residual taste heterogeneity, complicating identification and we cannot interpret the estimated coefficients of non-price variables as WTP estimates.

Equation (5) can be parameterized to become the GMNL model in WTP space by normalizing the parameter for the price variable to one inside the bracket, which results in:

$$\boldsymbol{\beta}_n = \sigma_n \alpha_n \begin{pmatrix} 1 \\ \frac{1}{\alpha_n}(\boldsymbol{\beta}^{np} + \boldsymbol{\Gamma}^{np}\mathbf{w}_n^{np}) \end{pmatrix} = \sigma_n \alpha_n \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{np} + \boldsymbol{\Pi}_c^{np}\mathbf{w}_n^{np} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{\beta}^{np}$, $\boldsymbol{\Gamma}^{np}$, and $\mathbf{w}_n^{np}$ are $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, and $\mathbf{w}_n$ excluding the price parameter σ_n Furthermore, $\boldsymbol{\theta}^{np} = \boldsymbol{\beta}^{np}/\alpha_n$ and $\boldsymbol{\Pi}_c^{np} = \boldsymbol{\Gamma}^{np}/\alpha_n$. As pointed out by Train and Weeks [36], the common denominators in $\boldsymbol{\theta}^{np}$ and $\boldsymbol{\Pi}_c^{np}$ induce correlation among all non-price parameters.

Similarly, we can parameterize the RPL model in preference space in equation (3) into the RPL model in WTP space by normalizing the parameter for the price variable, which results in:

$$\boldsymbol{\beta}_n = \alpha_n \begin{pmatrix} 1 \\ \frac{1}{\alpha_n}(\boldsymbol{\beta}^{np} + \boldsymbol{\Gamma}^{np}\mathbf{w}_n^{np}) \end{pmatrix} = \alpha_n \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{np} + \boldsymbol{\Pi}_c^{np}\mathbf{w}_n^{np} \end{pmatrix}. \tag{7}$$

The only difference between the GMNL model in WTP space and the RPL model in WTP space is that the former model includes an additional multiplier, the scale heterogeneity parameter σ_n.

3.2.3. Specification and estimation

The price parameter α_n can either be a non-random or random parameter, and we specified α_n as a random parameter to allow consumers to have different sensitivity to price changes [6]. A normal distribution is commonly assumed for random parameters, but for the price parameter this assumption has been criticized [6,35,60]. A normally distributed price parameter implies that there is a segment of consumers who receive positive utility from increasing prices, which is usually not behaviorally plausible.

We assume that α_n follows a one-sided triangular distribution [6,9,61–63].[2] All non-price parameters are assumed to follow normal distributions. In all four models, the parameters are specified to be uncorrelated. In the preference space models, this implies that, conditional on the mean vector and the chosen marginal distributions, the individual-specific taste parameters for the non-price attributes are independent. In the WTP space models, however, this independence in the underlying preference space representation does not translate into independence of the WTP coefficients. Because each WTP parameter is expressed as a ratio involving the random price coefficient, the normalized price parameter acts as a common denominator of the price parameter α_n in the WTP space. As a result, all non-price parameters will be correlated [36]. This structure implies that any individual who is more or less price sensitive will tend to have all of their WTP terms scaled up or down together, which can be viewed as a behaviorally plausible link between tradeoffs across attributes and overall price sensitivity. The WTP estimates from the models in preference space are obtained by using the Delta method [38].

We estimate all models by maximum simulated likelihood using NLOGIT 5 [38,65]. We specify panel data structure in all models. Train [31] found that 100 Halton draws outperformed 1000 random draws. In practice, 500 Halton draws is usually considered sufficient. We have tried different numbers of Halton draws, ranging from 100 to 2000. We find the model results stable in terms of log-likelihood function values, pseudo McFadden R^2 values, and estimated WTP distributions. In this paper, we report the estimation results with 500 Halton draws. We specify that a model converged, when the gradient and the change in any parameter value each is smaller than 0.00001.

4. Results and discussion

Table 2 reports estimation results of the GMNL and RPL models in preference and WTP space. The McFadden Pseudo R^2 values were above 0.4 in all models [66], which indicates a good fit [67, p. 54].

From our results, the GMNL models have better fit than the RPL models but the improvements are marginal. In preference space, the GMNL model improves the LLFVs as compared with the RPL model by 35 points (from −2028.59 to −1993.60). The improvement is 0.01 in terms of the McFadden Pseudo R^2 value. In WTP space, the improvements are even smaller. The GMNL model improves the LLFV of the RPL model by 4 points (from −2096.60 to −2093.08) and the improvement in term of McFadden Pseudo R^2 is less than 0.01 (from 0.5851 to 0.5858). These conclusions do not change when alternative model fit statistics, for example Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) are used.

These marginal improvements in model fit of the GMNL model as compared with the RPL model in preference space, is consistent with previous findings. Hildebrand et al. [23] found that the GMNL model only improved LLFV by 2 points over the RPL model. Fiebig et al. [1] estimated the RPL and GMNL models in preference space using nine datasets. The GMNL model had superior LLFVs as compared with the RPL model in all nine datasets, but only for two datasets the GMNL model improved the LLFVs by more than 100 points. Greene and Hensher [2] found that the GMNL model in preference space only improved LLFV by about 45 points as compared with the RPL model in preference space. Czajkowski, Giergiczny and Greene [12] found that when the parameters were uncorrelated, the GMNL model in preference space improves the LLFV by 559 points as compared with the RPL model in preference space. However, with the same dataset, the performance of the RPL model in preference space improved greatly when we allowed for correlation among the parameters, and the difference in LLFVs was reduced to 101 points. Moreover, these four cases, in which the GMNL model improved LLFVs by more than 100 points, do not necessarily suggest that the GMNL model sometimes largely improves LLFVs as compared to the RPL model in preference space. In all four cases, the base LLFVs were large and the improvements relative to the base LLFVs were small.[3]

When the RPL model was first introduced, it improved model fits greatly compared to the MNL model, which was the standard model back then. For example, compared to the MNL model, the RPL model in preference space improves LLFVs by over 12,000 points in Czajkowski, Giergiczny, and Greene's [12] study. In our results, the RPL model in preference space improves LLFV by more than 500 points as compared with the MNL model.[4] Therefore, synthesizing our results and the surveyed GMNL model studies, we conclude that comparing to the progress from the MNL model to the RPL model, which we believe was revolutionary, the progress from the RPL model to the GMNL is meaningful, important, and, at the same time, modest.

In both GMNL specifications, in preference and in WTP space, the estimated τ parameter is significantly different from zero, which confirms the presence of scale heterogeneity in our data. This means that, conditional on the observed attributes, respondents differ systematically in the dispersion of their unobserved utility. Some respondents exhibit more deterministic choice patterns, while others make noisier choices. This is consistent with the notion that individuals vary in their engagement with the choice experiment, decision certainty, or consistency in applying their preferences during the experiment.[5] The GMNL model in preference space achieves a better in-sample fit than the RPL model in preference space, with 35-point improvement in the log-likelihood and a 0.01 increase in McFadden's Pseudo R^2. The

[2] We have tried more common distributions before choosing one-sided triangular distribution. First, we used a lognormal distribution [64] for the price parameter. However, the GMNL model in preference space failed to converge, and the RPL model in preference converged but with an unstable standard error. We tried different numbers of draws per iteration, ranging from 100 to 2000 draws per iteration. The mean value and the standard error of the lognormal price parameter were still large and unstable. The standard error of the mean was about one million times larger than the standard errors of other mean values estimated by the model. Next, we estimated models with a price parameter that followed an unconstrained triangular distribution. However, in some cases the resulting triangular distribution included positive values in its distribution, which would be behaviorally implausible. By using a one-sided triangular distribution, which is a constrained version of the triangular distribution, we restrict the price parameter to be non-positive.

[3] The two datasets in Fiebig et al. [1], in which the GMNL model in preference space improves LLFVs by more than 100 points, were the Pizza B and the Holiday B datasets. For these two datasets, the LLFVs of the RPL models in preference space were over −5500 and −11,000, respectively. In Czajkowski, Giergiczny, and Greene [12], the LLFVs in the RPL model in preference space with uncorrelated parameters and correlated parameters were −17,613 and −17,055, respectively. These LLFVs are much larger than commonly seen. For example, the RPL models in preference space had LLFVs of around −1400 and around −2500 for the Pizza A and the Holiday A datasets in Fiebig et al. [1], respectively. When base LLFVs are large, it is natural that the improvements by the GMNL models in preference space are large in absolute values, but compared to the sizes of the base LLFVs, the improvements are still small.

[4] The LLFV of the standard MNL model is −2546.17. Because comparing the standard MNL is not part of the paper's main objectives, we do not report the detailed results of the MNL model in this paper.

[5] We do not attempt in this paper to relate the estimated scale heterogeneity to respondent characteristics. In principle, one could construct individual-specific scale measures from the GMNL models and examine how these relate to demographics or attitudinal variables from the post-experiment questionnaire, for example, to test whether more engaged or experienced fish consumers exhibit more deterministic choices. Such an analysis would be informative but would extend beyond our primary objective. We leave a detailed investigation of the determinants of scale heterogeneity to future work.

Table 2
Estimation results.

Attribute	Preference Space				WTP Space			
	GMNL		RPL		GMNL		RPL	
	Coef.	SD	Coef.	SD	Coef.	SD	Coef.	SD
Price in € per kilogram	−0.40***	0.40***	−0.31***	0.31***	−0.47***	0.47***	−0.34***	0.34***
	(0.03)	(0.03)	(0.01)	(0.01)	(0.04)	(0.04)	(0.02)	(0.02)
Wild cod	7.36***	1.34***	5.91***	1.36***	16.43***	7.28***	17.81***	8.44***
	(0.54)	(0.20)	(0.31)	(0.21)	(0.73)	(0.57)	(0.69)	(0.59)
Farmed cod	7.11***	1.09***	5.06***	1.07***	16.45***	5.87***	16.16***	6.25***
	(0.51)	(0.22)	(0.29)	(0.19)	(0.46)	(0.47)	(0.53)	(0.67)
Farmed salmon	7.26***	1.88***	5.25***	2.15***	18.62***	6.94***	16.71***	6.42***
	(0.53)	(0.16)	(0.27)	(0.21)	(0.43)	(0.42)	(0.42)	(0.43)
Wild monkfish	8.52***	0.93***	6.58***	0.95***	19.49***	11.31***	20.26***	11.41***
	(0.61)	(0.29)	(0.32)	(0.25)	(0.82)	(1.10)	(0.84)	(0.94)
Farmed pangasius	−0.09	3.41***	−0.04	3.29***	−3.85*	13.18***	−0.91	10.22***
	(0.57)	(0.50)	(0.52)	(0.51)	(2.16)	(2.05)	(1.74)	(1.63)
MSC label	0.57***	1.17***	0.43**	0.91***	1.89***	4.18***	2.26***	3.13***
	(0.22)	(0.23)	(0.18)	(0.25)	(0.64)	(0.74)	(0.61)	(0.84)
Organic AB label	0.70***	0.87***	0.76***	0.59***	1.92***	1.70***	2.22***	2.67***
	(0.14)	(0.16)	(0.12)	(0.19)	(0.31)	(0.22)	(0.45)	(0.47)
τ	0.67***		<u>0</u>		0.62***		<u>0</u>	
	(0.06)				(0.08)			
σ_n	1.00	0.72	<u>1</u>		1.00	0.66	<u>1</u>	
γ	0.67***		<u>0</u>		<u>0</u>		<u>0</u>	
	(0.10)							
Log likelihood function	−1993.60		−2028.59		−2093.08		−2096.60	
Akaike Information Criterion	4021.20		4087.20		4218.20		4223.20	
Bayes Information Criterion	4118.80		4173.30		4310.00		4309.30	
McFadden pseudo R^2	0.61		0.60		0.59		0.59	

A parameter that is significant at the 0.10, 0.05, and 0.01 level of significance is marked with *, **, and ***, respectively.
Notes.
The numbers in the parentheses are the standard errors. GMNL and RPL denote the generalized multinomial logit and the random parameter logit model, respectively. All non-price attributes are dummy variables. Underlined numbers are fixed parameters and were not estimated. Each price parameter was assumed to follow a one-sided triangular distribution. For the price parameters, which follow the one-sided triangular distribution, the numbers in the SD columns are the limits of this distribution, where mode is zero and the estimated values are the upper limits.

improvement is likely due to the fact that the GMNL model captures scale heterogeneity.

In the GMNL model in preference space, γ is estimated to be 0.67 and is statistically significant at 0.0.1 level. This means that the GMNL model in preference space finds statistical significant interaction between scale heterogeneity and residual taste heterogeneity. This interaction is assumed away when estimating the GMNL model in WTP space, which partially explains the reduction in model fit when comparing with the GMNL model in preference space. The RPL models, in preference space and in WTP space, do not contain a parameter like γ, which also partially explains the better model fit of the GMNL model in preference space as compared to the RPL models, which is reflected in LLFVs.

Estimating the GMNL model in WTP space reduces model fit to the data as compared to estimation in preference space. In our study, the LLFV is reduced by 99 points and McFadden Pseudo R^2 is reduced by 0.02. These reductions in model fit when estimating the GMNL model in WTP space as compared to in preference space, are consistent with previous results [9,19]. Hensher and Greene [9] reported a reduction of 71 points in LLFV and a reduction of 0.02 in McFadden Pseudo R^2. Michaels-Igbokwe et al. [19] did not report McFadden Pseudo R^2 values but found a reduction of 53 points in LLFV. We have not found a case in the literature that the GMNL model estimation in WTP space increases model fit, when compared to the GMNL model in preference space.

For the RPL models in our study, the model fit is also reduced in WTP space as compared to the preference space. This finding is consistent with some previous findings (e.g., Ref. [36,37,39]), but not all. For example, Train and Weeks [36] found that when parameters are specified to be uncorrelated, the LLFV was reduced by 64 points when the RPL model was estimated in WTP space compared to when estimated in preference space. Moreover, they found that when parameters were specified to be correlated, the reduction due to estimation in WTP space was slightly less, at 49 points. However, we have also found studies where estimating the RPL model in WTP space improved LLFV. Scarpa, Thiene, and Train [33] found that the LLFV were improved by estimating RPL model in WTP space by 303 points and by 58 points, when the parameters were specified to be uncorrelated and correlated, respectively. Scarpa, Thiene and Hensher [68] and Gu, Norman and Viney [69] also found that the RPL model in WTP space fitted the data better than the RPL model in the preference space.

The reduction in model fit of GMNL and RPL models in WTP space in our results likely come from three sources. First, WTP space estimations of the GMNL model and the RPL model impose additional structure relative to their preference space counterparts. Normalizing the price coefficient to the scale parameter and interpreting the remaining coefficients as WTP introduces nonlinearity in the random parameters and mechanically induces correlation among all non-price coefficients through the common denominator. Second, the one-sided triangular distribution for the price parameter constrains the values of price coefficient distribution to be non-positive. This avoids behaviorally implausible positive price coefficients, but it also imposes a relatively rigid functional form on the distribution of price sensitivity. Third, for the GMNL model in WTP space, fixing $\gamma = 0$ further restricts how scale heterogeneity can interact with preference heterogeneity. As a result, the WTP-space models are more constrained in the way they can represent the joint distribution of taste heterogeneity and scale heterogeneity, which likely has reduced their ability to match certain patterns in the choice data compared with the corresponding models in preference space.

The estimated WTP values are presented in Table 3. The mean WTP values and the Confidence Intervals (CIs) of the two models estimated in WTP space are directly given by the parameter estimates. The WTP values and the CIs of the two models in preference space are calculated by using the Delta method [70] using equation (4).

No model has consistently higher or lower WTP estimates than the other models. The estimated WTP per kilogram of wild cod, farmed cod, farmed salmon, and wild monkfish are €16.43, €16.45, €18.62, and €19.49 according to the GMNL model in WTP space. For these four fish types the mean WTP estimates are significantly different from zero at the 1% level. For wild cod and farmed salmon, which are the most commonly purchased types in France, the estimated prices are within the range of prices charged by local stores at the time of the experiment. Back then, farmed cod was not available in local stores, and monkfish was sold in small quantities at market prices significantly above our WTP estimates. Except for pangasius, the mean WTP values do not differ much across models.

A dataset of 2300 choices made by a sample of 144 participants is acceptable but modest for complex MMNL models. Without sufficient variation in choices and in attributes, instability in estimates may occur. Our results across four different models are largely stable, meaning that the choice dataset contains sufficient variations such that the GMNL models and the RPL models in preference space and in WTP space can identify stable values across most of the parameters, with the exception of the coefficient for pangasius. This assures the quality of our empirical dataset.

Because the data is from 2008, the data and the results from the data therefore reflect consumer preferences and market conditions as of 2008. Since then, ecolabel awareness, trust, and market penetration in seafood markets have evolved. Consequently, the absolute WTP values from our results should be interpreted as specific for around 2008. The WTP values likely do not reflect the current market situations in France.

Our estimated negative mean WTPs for farmed pangasius are outside the price range used in the experiment, which indicates that less than 50% of respondents were willing to pay anything for it. The 95% CIs of WTP for pangasius overlap across models. However, while the GMNL model in WTP space identifies the mean WTP for pangasius that is significant at 10% level, none of the other three models finds statistically significant mean WTP for pangasius. This also means that the mean WTP for pangasius is not statistically significant at 5% level or at 1% level in all of the models. This discrepancy suggests instability in estimates for the preference for pangasius.

Several factors may have contributed to the instability. Pangasius was chosen in 10% of the choice sets where it was included even though it was the cheapest alternative.[6] Because the choice experiment was hypothetical, the participants didn't need to consider their budget constraints and didn't have to pay for what they chose. In real life, with a budget constraint, a cheap alternative like pangasius is likely chosen more times among consumers that are not very fond of it, due to its price, especially for consumers under tight food budget, than in a stated choice experiment without an actual budget constraint. This means that we potentially have a hypothetical bias that has amplified the lack of desirability of pangasius and has negatively affected the choice probability of selecting pangasius. This can explain why we find that many participants were reluctant to choose pangasius as an inexpensive substitute, despite its price advantage.

Our results suggest that pangasius occupies a marginal market position in this experiment that goes beyond its low price. The negative and statistically insignificant mean WTP, combined with the fact that pangasius is chosen relatively rarely even though it is the cheapest option, indicates that a substantial share of respondents attaches very low or even negative utility to this product at any positive price. This suggests that pangasius is perceived as an inferior or low desirability alternative, possibly due to unfamiliarity, country-of-origin perceptions, and/or concerns about quality or production practices, in addition to the potential hypothetical bias, although our data do not allow us to verify whether or which of these mechanisms were the underlying cause(s).

The mean WTP estimates for the two labels are significantly different from zero at the 5% level of significance. The WTP for the MSC label ranges from €1.39 to €2.26 in the four models, and the WTP for the AB label ranges from €1.73 to €2.44. This corresponds to a premium of about 10%, which is consistent with previous estimates [45,55,71]. The slightly higher WTP values for the AB label may indicate that consumers have more trust in this label.

The estimated mean WTP values from the GMNL model are within the 95% CIs of the corresponding WTP values for the RPL model, and vice versa, with one exception.[7] The CIs for farmed salmon from the GMNL model in WTP space and the RPL model in WTP space are close but do not overlap. This indicates that the combination of scale heterogeneity, which is captured by the GMNL model, and the WTP space parameterization have a small but statistically discernible impact on the estimated WTP for farmed salmon. On one hand, consumer preference for farmed salmon may be very different. Some may like farmed salmon a lot, while others strongly dislike farmed salmon and strongly prefer wild salmon. On the other hand, farmed salmon was probably one of the most familiar fish for the participants. Therefore, participants might have used very different decision heuristics when facing a choice set that contained farmed salmon. In such a case, differences in how the models model and capture different types of heterogeneities, such as inclusion of scale heterogeneity in the GMNL models and as the induced correlation structure in models estimated in WTP space, could have had a larger influence on the estimated mean WTP for farmed salmon than for other fish types.

We also examined the distribution of the WTP estimates for the individual participants. Fig. 1 shows the Kernel Distribution Estimates (KDEs) for the WTP for the fish types and the labels from the GMNL and the RPL models. The KDEs for one attribute's WTP estimates are similar across the four models, which is not surprising given that most of the CIs overlap.

This means that despite the presence of scale heterogeneity in the dataset, the WTP estimates by accounting or not accounting scale heterogeneity do not statistically differ for the majority of the WTP

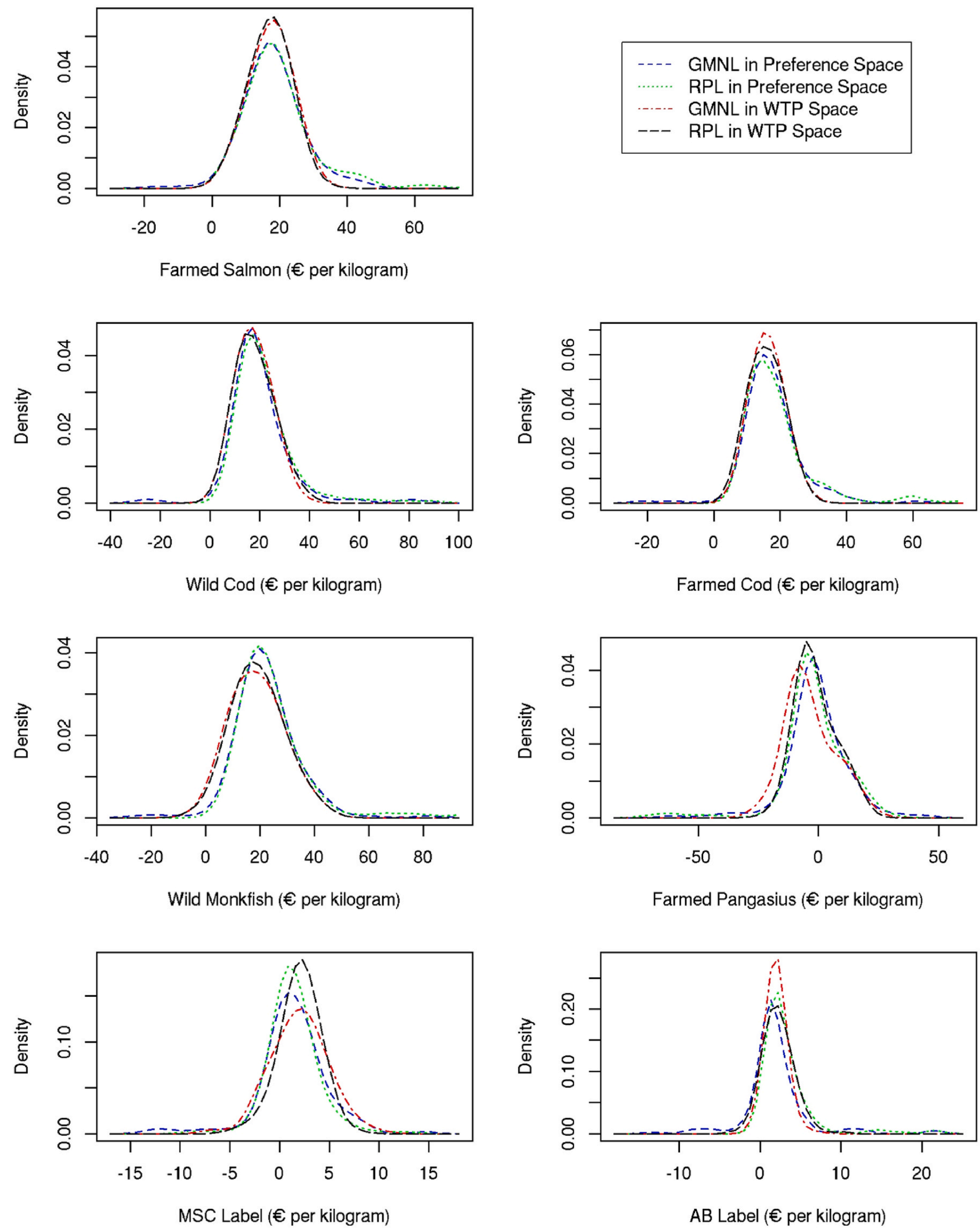
Table 3
Mean WTP estimates and 95% confidence intervals (in € per kilogram).

Attribute	Preference Space		WTP Space	
	GMNL	RPL	GMNL	RPL
Wild cod	18.22***	19.08***	16.43***	17.81***
	[17.18,19.26]	[17.71,20.45]	[15.00,17.86]	[16.46,19.16]
Farmed cod	17.58***	16.34***	16.45***	16.16***
	[16.70,18.46]	[15.04,17.64]	[15.54,17.35]	[15.11,17.21]
Farmed salmon	17.95***	16.96***	18.62***	16.71***
	[17.00,18.90]	[15.64,18.28]	[17.78,19.46]	[15.89,17.52]
Wild monkfish	21.09***	21.24***	19.49***	20.26***
	[20.04,22.13]	[19.86,22.62]	[17.89,21.10]	[18.61,21.90]
Farmed pangasius	−0.22	−0.13	−3.85*	−0.91
	[-2.98,2.53]	[-3.42,3.15]	[-8.08,0.38]	[-4.33,2.51]
MSC label	1.42***	1.39**	1.89***	2.26***
	[0.37,2.46]	[0.24,2.54]	[0.64,3.14]	[1.07,3.46]
Organic AB label	1.73***	2.44***	1.92***	2.22***
	[1.05,2.41]	[1.70,3.18]	[1.31,2.53]	[1.33,3.11]

Notes.
A parameter that is significant at the 0.10, 0.05, and 0.01 level of significance is marked with *, **, and ***, respectively. The numbers in the brackets are the 95% confidence intervals. GMNL and RPL denote the generalized multinomial logit and the random parameter logit models, respectively.

[6] Pangasius appeared as an alternative in 816 choice situations. It was chosen only 82 times.

[7] Similar CIs were obtained based on the Krinsky–Robb method [72,73].



Note: Each plot is a comparison of the KDEs for an attribute from the RPL and the GMNL models.

Fig. 1. Distributions of WTP for Fish Types and Labels (in € per kilogram).

estimates, with the exception of 95% WTP CIs for farmed salmon from the GMNL model in WTP space and the RPL model in WTP space.

The induced correlation among WTP coefficients in the WTP space models has a behavioral interpretation. Because all non-price WTP terms share the normalized price coefficient in the denominator, individuals who are more price-sensitive will tend to have lower WTP values for all attributes, while less price-sensitive individuals will tend to have higher WTP values across the board. This creates a coherent pattern in which tradeoffs between attributes are linked through a common underlying price sensitivity. In contrast, the preference space models with independent random coefficients allow each attribute's utility weight to vary independently, so that an individual could, in principle, have a high utility weight for one attribute and a low weight for another even if their overall price sensitivity would suggest otherwise.

In our data, however, the kernel density estimates in Fig. 1 and the confidence intervals in Table 3 suggest that this implicit correlation structure in estimation in WTP space does not lead to substantially different WTP distributions or mean WTP estimates compared with estimation in preference space. The main features of the WTP distributions, which include dispersion in the empirically relevant range and relative ranking across fish types and labels, are very similar across specifications. This indicates that, for this dataset, the additional behavioral structure implied by the induced correlation in WTP space does not change the WTP estimates in statistical or economic meaningful ways, even though it represents a conceptually appealing way to link attribute-specific tradeoffs through price sensitivity.

Several KDEs of the WTP exhibit fat tails. Although the KDEs largely overlap, we can still examine the changes in distributional shapes. One important theoretical advantage of estimation in WTP space is that it avoids potential problems of fat tails in WTP distribution from estimation in preference space [35,36]. Both the GMNL model and the RPL model in preference space, in blue and in green, respectively, produce WTP KDEs for farmed cod with fat right tails (Fig. 1). Both models in preference space produce prominent fat right tails in the WTP KDEs for farmed salmon. Moreover, both models in preference space produce WTP KDEs for wild cod that have slight fat right tails.

However, estimation in WTP space does not always guarantee that fat tails will not occur. Take WTP KDEs for wild monkfish as an example. Both the GMNL model and the RPL model in WTP space produce fat left tails when compared to their counterparts in preference space (Fig. 1). The WTP KDEs for the ecolabels are more nuanced. For WTP KDEs for the MSC label, the GMNL model in preference space and the RPL model in WTP space produce slight fat right tails, while the GMNL model in WTP space produces a fat left tail in the WTP KDE. Comparing WTP KDEs for the AB label, the RPL models in both spaces produce fat right tails, while the GMNL model in preference produces a fat left tail. Because the WTP estimates for pangasius is unstable, we note that the instability is also reflected in KDEs shown in Fig. 1.

Overall, we have seen evidences that both support and contradict the theoretical advantage of having finite moments of WTP distributions through estimation in WTP space. There are more cases where estimation in WTP space produces thinner tails in Fig. 1. However, we cannot conclude that estimation in WTP space always guarantees thin tails. Our empirical results therefore suggest that the theoretical advantages of WTP space estimation do not always automatically translate into clear empirical gains in all applied settings.

Our results have practical implication. For datasets where estimation in preference space yields implausibly high WTP estimates such as large fat tails, we recommend estimation in WTP space as the technique that can potentially uncover WTP estimates that are closer to their true values.

Estimation in WTP space places restriction on the model, by normalizing the price parameter to one, such that often a model estimated in WTP space fits the data somewhat less well as the comparable model estimated in preference space. This means that if WTP estimation is not the only goal, estimation in preference space may be preferred, for example, if purchase behavior prediction or market share forecasting is important, because a model that fits data better will likely produce more accurate predictions and forecasting.

If possible, we recommend that a model is estimated in both preference space and in WTP space, so that the estimated coefficients, WTP estimates, model fit, and forecasting results can be compared. Through careful evaluation, we may choose estimation in preference space or in WTP space.

Taken the results of comparisons together, we suggest a decision framework for choosing between the GMNL model and the RPL model in applied work. When the primary objective is to obtain robust mean (or median) WTP estimates for policy analysis or market studies and there is no strong prior to suspect substantial scale heterogeneity, a standard RPL model in preference space is typically sufficient, because the RPL model is easier to estimate, widely implemented in standard software, and, in our data, yields WTP estimates that are very close to those from GMNL. By contrast, the GMNL model becomes attractive when (i) differences in scale across individuals or choice situations are of substantial interest, (ii) datasets combine very heterogeneous respondents or experimental conditions where error variance plausibly varies, and/or (iii) model comparison statistics conclude a large improvement in model fit by allowing for scale heterogeneity. In those cases, the extra complexity of the GMNL model is justified. We therefore recommend that applied researchers start with a RPL model in practice, then estimate a GMNL model as a robustness check. If the GMNL model does not materially change WTP estimates or significantly improve fit, the simpler RPL model can be the main specification, with the GMNL model relegated to robustness analysis. The GMNL model remains attractive in three cases: (1) when the explicit modeling of scale heterogeneity is of substantive interest, (2) when the focus is on describing choice probabilities as accurately as possible within sample, and (3) when the two models produce quite different results and we have reason to believe that the results from the GMNL model are closer to their true values, e.g. due to value(s) from test statistic(s) or due to a theoretical prior.

In the end of this section, we would like to discuss the data limitation of this paper and the implications. An important limitation of our study is the age of the data. The stated choice experiment was conducted in December 2008.The French seafood market has changed since then. Since 2008, ecolabel awareness, the range and visibility of ecolabels, and consumer preference for ecolabels and organic products have all evolved over time. As a result, the absolute WTP levels we report for ecolabeled and organic fish should be interpreted as describing preferences in the period around 2008, not as current estimates for today's consumers.

A further implication of the age of the data is that the direct policy relevance of the WTP estimates is limited. The absolute WTP values and premiums we report should not be directly used to infer policy or used for cost–benefit analysis, without additional updated data.

Our main conclusions are therefore methodological. They concern the relative performance and robustness of different MMNL specifications for WTP estimation, which are not affected by the age of the dataset. The family of multinomial logit models is widely used both in research and in application. The comparison of the GMNL and the RPL models in preference space and in WTP space, presented in this study, is interesting and relevant for the literature of DCM. Although this study contains quite technical details, due to the nature of the study when comparing econometric models and their results, this study nevertheless sheds lights for practitioners, for example, business owners and analysts.

5. Conclusions

A stated choice experiment including ecolabeled and unlabeled farmed and wild fish was conducted in France. We estimated the RPL and GMNL models in preference and in WTP space. French consumers were willing to pay significantly more for ecolabeled wild fish and organically labeled farmed salmon and cod than for their unlabeled counterparts. On average, both labels attracted a premium of about 10% in 2008.

We emphasize that the policy implications of our WTP estimates are constrained by the age of the data. The experiment reflects market conditions and ecolabel availability in 2008, and the absolute premiums we estimate for MSC and organic labels should be viewed as historical and context-specific values. They are not intended to be used directly as current policy inputs without new data.

Our principal contribution is methodological. We demonstrate that alternative MMNL specifications, the GMNL model and the RPL model in preference space and in WTP space, deliver very similar WTP estimates. This suggests that a parsimonious[8] RPL model in preference space is adequate for WTP estimation for our dataset. These methodological insights are not affected by the age of our dataset.

Although all four models fit the data well, the GMNL model in preference space, which incorporates scale heterogeneity and taste heterogeneity and allows these two sources of heterogeneity to interact, statistically fits slightly better than the other three models. However, this improved fit has negligible consequences for the WTP estimates in our study. The WTP estimates from the four models are similar for four of the five fish types. The four models also produce quite similar WTP estimates for the two labels included in the experiment. The estimated WTP values from one model are within the 95% CIs of the corresponding WTP values for the three other models, and vice versa, with only one exception. Moreover, all four models produce similar KDE plots of individual-specific WTP estimates across participants. This means that the existence of statistically significant scale heterogeneity in GNML models did not change the WTP estimates in an economically or statistically significant ways, as compared to the WTP estimates from RPL models.

Our results do not imply that GMNL is unnecessary. While the practical difference in WTP estimates may be negligible, the statistical evidence favors GMNL, reflecting GMNL's additional flexibility in capturing scale heterogeneity and accounting for the interaction between taste heterogeneity and scale heterogeneity. However, this improvement in in-sample fit is modest relative to the gains achieved as compared to, for example, the improvement of the RPL model over the MNL model.

The model choice insights from this paper will be helpful to both academic researchers and practitioners in business, government, and other organizations. Our results imply that a standard RPL model in preference space will likely produce similar WTP estimates as the RPL model in WTP space and as the GMNL model in preference or in WTP space. Considering the philosophy of model parsimony, we shall choose the model that has fewer parameters and simpler structure. A parsimonious model has less risk of overfitting because it has fewer parameters. Having fewer parameters also likely leads to fewer convergence problems, which means higher tolerance of sample size and data quality. A parsimonious model tends to have better out-of-sample predicting powder than a comparable model with more parameters. This means that if a RPL model and a GMNL model produce comparative WTP estimates from a dataset, the choice lies in the hands of the modeler, because the GMNL model likely will have a small but better statistical fit and arguably provide slightly more accurate WTP estimates. On the other hand, the RPL model may be preferred due to parsimony and potentially higher predictive power out of sample.

The GMNL model remains attractive when the explicit modeling of scale heterogeneity is of substantial interest, or when the focus is on describing choice probabilities as accurately as possible within sample. If a RPL model and a GMNL model produce quite different WTP estimates from a data set and the GMNL model has a statistical better fit, we encourage researchers and practitioners to choose the GMNL model, because the better fit and the differences in WTP estimates are likely due to pronounced scale heterogeneity, which the RPL model cannot capture.

Because a GMNL model has more parameters than its counterpart RPL model, for a dataset that has smaller sample size or other reasons that resulted insufficient variation in the dataset itself, a situation may arise that one can estimate the RPL model but not the GMNL model with the dataset. In such case, one will have to use the RPL model, which also reflects the parsimony of the RPL model as compared to the GMNL model.

Our empirical results of KDEs of the WTP estimates do not conclusively support the theoretical advantage of having finite moments in WTP estimates by estimating in WTP space. We have more evidence that supports this theoretical advantage. We also have evidences where estimation in WTP space produces fat tails.

Our comparison of model fit reduction of the GMNL model in WTP space is consistent with the previous findings. Intriguingly, the empirical results of ours and from the literature find that, unlike the RPL model, the GMNL model always produce lower model fit when estimated in WTP space, as compared to the GMNL model in preference space. We do not know if this observation will hold, when more datasets are estimated in the GMNL model in preference space and in WTP space. If this holds, it will be interesting to explore the reasons and the mechanisms behind it.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

The Research Council of Norway, grants 178300/I10 and 199564/I10, provided financial support for this research.

Declaration of competing interest

The author declares that he does not have known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Sylvie Issanchou and Géraldine Enderli at l'institut national de la recherche agronomique's (INRA) laboratory (flavour, vision, consumer behaviour) in Dijon and Pierre Combris at INRA in Paris helped us with data collection. Their help and their valuable comments are highly appreciated. I'm in debted to Frode Alfnes and Kyrre Rickertsen, for their guidance on my work on this paper during my time as their PhD student. I thank Gerald E. Shively, John Rose, Mikołaj Czajkowski, William Greene, Søren Bøye Olsen, Geir Sogn-Grundvåg, Knut Einar Rosendahl, Igor Pipkin, Cathy A. Roheim, Richard B. Howarth, and Anne Aitken for their comments on earlier drafts of the paper. The constructive comments and suggestions from the three anonymous reviewers greatly helped improve the manuscript.

Data availability

The author does not have the permission to share data.

References

[1] D.G. Fiebig, M.P. Keane, J. Louviere, N. Wasi, The generalized multinomial logit model: accounting for scale and coefficient heterogeneity, Mark. Sci. 29 (3) (2010) 393–421.

[2] W.H. Greene, D.A. Hensher, Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models, Transportation 37 (3) (2010) 413–428.

[3] D. Revelt, K.E. Train, Mixed logit with repeated choices: Households' choices of appliance efficiency level, Rev. Econ. Stat. 80 (4) (1998) 647–657.

[4] D. McFadden, K.E. Train, Mixed MNL models for discrete response, J. Appl. Econom. 15 (5) (2000) 447–470.

[8] We consider model parsimony as the principle of modeling data using the simplest possible explanation, with the fewest independent variables or parameters, that still provides a good fit. We note that some scholars in the literature adopt a different definition of model parsimony, such as [59], who considered the GMNL model more parsimonious than the RPL model.

[5] D. McFadden, Economic choices, Am. Econ. Rev. 91 (3) (2001) 351–378.
[6] D.A. Hensher, W.H. Greene, The mixed logit model: the state of practice, Transportation 30 (2) (2003) 133–176.
[7] S. Hess, J.M. Rose, Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39 (6) (2012) 1225–1239.
[8] S. Hess, K. Train, Correlation and scale in mixed logit models, Journal of Choice Modelling 23 (2017) 1–8.
[9] D.A. Hensher, W.H. Greene, Valuation of travel time savings in WTP and preference space in the presence of taste and scale heterogeneity, J. Transport Econ. Pol. 45 (3) (2011) 505–525.
[10] D.A. Hensher, J.M. Rose, Z. Li, Does the choice model method and/or the data matter? Transportation 39 (2) (2012) 351–385.
[11] J. Rose, S. Hess, W.H. Greene, D.A. Hensher, The generalizied multinomial logit model: misinterpreting scale and preference heterogeneity in discrete choice models or untangling the un-untanglable?, in: 13th International Conference on Travel Behaviour Research, 2012. Toront.
[12] M. Czajkowski, M. Giergiczny, W.H. Greene, Learning and fatigue effects revisited: investigating the effects of accounting for unobservable preference and scale heterogeneity, Land Econ. 90 (2) (2014) 324–351.
[13] M. Czajkowski, T. Kądziela, N. Hanley, We want to sort! assessing households' preferences for sorting waste, Resour. Energy Econ. 36 (1) (2014) 290–306.
[14] X. Li, C. Clark, K. Jensen, S. Yen, Will consumers follow climate leaders? The effect of manufacturer participation in a voluntary environmental program on consumer preferences, Environ. Econ. Pol. Stud. 16 (1) (2014) 69–87.
[15] M. Czajkowski, N. Hanley, J. LaRiviere, The effects of experience on preferences: theory and empirics for environmental public goods, Am. J. Agric. Econ. 97 (1) (2015), 333-315.
[16] M.M. Rahman, A.K. Bohara, Assessing the preference and spatial dependence of a solid waste management system in Nepal: a choice experiment approach, J. Environ. Manag. 327 (2023) 116805.
[17] S.A. Knox, R.C. Viney, Y. Gu, A.R. Hole, D.G. Fiebig, D.J. Street, M.R. Haas, E. Weisberg, D. Bateson, The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products, Soc. Sci. Med. 83 (2013) 70–80.
[18] L.B. Pedersen, J. Riise, A.R. Hole, D. Gyrd-Hansen, GPS' shifting agencies in choice of treatment, Appl. Econ. 46 (7) (2014) 750–761.
[19] C. Michaels-Igbokwe, M. Lagarde, J. Cairns, F. Terris-Prestholt, Designing a package of sexual and reproductive health and HIV outreach services to meet the heterogeneous preferences of young people in Malawi: results from a discrete choice experiment, Health Economics Review 5 (1) (2015) 1–19.
[20] J. Pancras, D.K. Dey, A comparison of generalized multinomial logit and latent class approaches to studying consumer heterogeneity with some extensions of the generalized multinomial logit model, Appl. Stoch Model Bus. Ind. 27 (6) (2011) 567–578.
[21] T. de-Magistris, A. Gracia, Do consumers care about organic and distance labels? An empirical analysis in Spain, Int. J. Consum. Stud. 38 (6) (2014) 660–669.
[22] C. Escobar, Z. Kallas, J.M. Gil, Impact of the economic and the political changes on consumers' wine preferences in Catalonia (Spain): a generalized multinomial logit approach, in: 143rd Joint EAAE/AAEA Seminar, March 25-27, 2015, Naples, Italy. European Association of Agricultural Economists, 2015.
[23] K. Hildebrand, C. Chung, T.A. Boyer, M. Palma, Does change in respondents' attention affect willingness to accept estimates from choice experiments? Appl. Econ. 55 (28) (2023) 3279–3295.
[24] K.H. Lim, W. Hu, Contextual reference price in choice experiments, American Journal of Agriculture Economics 105 (4) (2023) 1288–1306.
[25] V.H. Dang, T.V. Doan, Consumer preferences for country of origin of foreign fruit: does consumer animosity matter? Cogent Bus. Manag. 11 (1) (2024) 2330715.
[26] A.D. Murray, R.K. Gallardo, A. Prakash, Mexican consumers' attitudes toward irradiated and imported apples, Agribusiness 41 (3) (2025) 694–718.
[27] R.K. Gallardo, X. Ma, S.P. Galinato, T. Northfield, Unlocking consumer preferences: estimating the willingness to pay for integrated pest management practices, Journal of the Agricultural and Applied Economics Association 4 (4) (2025) 483–506.
[28] D. Doiron, J. Hall, P. Kenny, D.J. Street, Job preferences of students and new graduates in nursing, Appl. Econ. 46 (9) (2014) 924–939.
[29] Z. Ding, X. Fan, J. Zhang, Self-employed versus paid-employed: what are the different preferences for microfinance? Experimental evidence from rural China, Sage Open 14 (2) (2024) 21582440241239807.
[30] I. Mauerer, G. Tutz, Heterogeneity in general multinomial choice models, Stat. Methods Appl. 32 (1) (2023) 129–148.
[31] K.E. Train, Discrete Choice Methods with Simulation, Second Edition, Cambridge University Press, New York, NY, 2009.
[32] F. Alfnes, A.G. Guttormsen, G. Steine, K. Kolstad, Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives, Am. J. Agric. Econ. 88 (4) (2006) 1050–1061.
[33] R. Scarpa, M. Thiene, K.E. Train, Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the alps, Am. J. Agric. Econ. 90 (4) (2008) 994–1010.
[34] M.C.J. Bliemer, J.M. Rose, Construction of experimental designs for mixed logit models allowing for correlation across choice observations, Transp. Res. Part B Methodol. 44 (6) (2010) 720–734.
[35] A. Daly, S. Hess, K.E. Train, Assuring finite moments for willingness to pay in random coefficient models, Transportation 39 (1) (2012) 19–31.
[36] K.E. Train, M. Weeks, Discrete choice models in preference space and willingness-to-pay space, in: R. Scarpa, A. Alberini (Eds.), Applications of Simulation Methods in Environmental and Resource Economics, Springer, The Netherlands, 2005, pp. 1–16.
[37] G. Sonnier, A. Ainslie, T. Otter, Heterogeneity distributions of willingness-to-pay in choice models, Quant. Market. Econ. 5 (3) (2007) 313–331.
[38] W.H. Greene, NLOGIT version 5 reference guide, Plainview, NY: Econometric Software (2012). Inc.
[39] A.R. Hole, J.R. Kolstad, Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment, Empir. Econ. 42 (2) (2012) 445–469.
[40] C.R. Wessells, J.G. Anderson, Consumer willingness to pay for seafood safety assurances, J. Consum. Aff. 29 (1) (1995) 85–107.
[41] M.F. Teisl, B. Roe, R.L. Hicks, Can eco-labels tune a market? Evidence from dolphin-safe labeling, J. Environ. Econ. Manag. 43 (3) (2002) 339–359.
[42] S. Jaffry, H. Pickering, Y. Ghulam, D. Whitmarsh, P. Wattage, Consumer choices for quality and sustainability labelled seafood products in the UK, Food Policy 29 (3) (2004) 215–228.
[43] W. Verbeke, F. Vanhonacker, I. Sioen, J. Van Camp, S. De Henauw, Perceived importance of sustainability and ethics related to fish: a consumer behavior perspective, AMBIO J. Hum. Environ. 36 (7) (2007) 580–585.
[44] D. Brécard, B. Hlaimi, S. Lucas, Y. Perraudeau, F. Salladarré, Determinants of demand for green products: an application to eco-label demand for fish in Europe, Ecol. Econ. 69 (1) (2009) 115–125.
[45] G. Sogn-Grundvåg, T.A. Larsen, J.A. Young, The value of line-caught and other attributes: an exploration of price premiums for chilled fish in UK supermarkets, Mar. Pol. 38 (2013) 41–44.
[46] H. Uchida, Y. Onozaka, T. Morita, S. Managi, Demand for ecolabeled seafood in the Japanese market: a conjoint analysis of the impact of information and interaction with other labels, Food Policy 44 (2014) 68–76.
[47] F. Asche, T.A. Larsen, M.D. Smith, G. Sogn-Grundvåg, J.A. Young, Pricing of eco-labels with retailer heterogeneity, Food Policy 53 (0) (2015) 82–93.
[48] J. Bronnmann, F. Asche, Sustainable seafood from aquaculture and wild fisheries: insights from a discrete choice experiment in Germany, Ecol. Econ. 142 (2017) 113–119.
[49] J. Weitzman, M. Bailey, Perceptions of aquaculture ecolabels: a multi-stakeholder approach in Nova Scotia, Canada, Mar. Pol. 87 (2018) 12–22.
[50] I. Ankamah-Yeboah, F. Asche, J. Bronnmann, M. Nielsen, R. Nielsen, Consumer preference heterogeneity and preference segmentation: the case of ecolabeled salmon in Danish retail sales, Mar. Resour. Econ. 35 (2) (2020) 159–176.
[51] L.H. Gulbrandsen, The emergence and effectiveness of the Marine Stewardship Council, Mar. Pol. 33 (4) (2009) 654–660.
[52] M. Thrane, F. Ziegler, U. Sonesson, Eco-labelling of wild-caught seafood products, J. Clean. Prod. 17 (3) (2009) 416–423.
[53] Marine Stewardship Council, Working Together for Sustainable Seafood. The Marine Stewardship Councilc Annual Report 2024-25 (2025). URL: https://www.msc.org/docs/default-source/default-document-library/about-the-msc/msc-annual-report-2024-2025.pdf.
[54] F. Alfnes, X. Chen, K. Rickertsen, Labeling farmed seafood: a review, Aquacult. Econ. Manag. 1 (2018) 1–26.
[55] C.A. Roheim, F. Asche, J.I. Santos, The elusive price premium for ecolabelled products: evidence from seafood in the UK market, J. Agric. Econ. 62 (3) (2011) 655–668.
[56] I. Olesen, A.I. Myhr, G.K. Rosendal, Sustainable aquaculture: are we getting there? Ethical perspectives on salmon farming, J. Agric. Environ. Ethics 24 (4) (2011) 381–408.
[57] C. Mauracher, T. Tempesta, D. Vecchiato, Consumer preferences regarding the introduction of new organic products. The case of the Mediterranean sea bass (Dicentrarchus labrax) in Italy, Appetite 63 (2013) 84–91.
[58] M. Keane, N. Wasi, Comparing alternative models of heterogeneity in consumer choice behavior, J. Appl. Econom. 28 (6) (2013) 1018–1045.
[59] Y. Gu, A.N. Hole, S. Knox, Fitting the generalized multinomial logit model in Stata, STATA J. 12 (2) (2013), 12(2), pp. 382–382-97.
[60] S. Hess, M. Bierlaire, J.W. Polak, Estimation of value of travel-time savings using mixed logit models, Transport. Res. Pol. Pract. 39 (2) (2005) 221–236.
[61] R. Brouwer, T. Dekker, J. Rolfe, J. Windle, Choice certainty and consistency in repeated choice experiments, Environ. Resour. Econ. 46 (1) (2010) 93–109.
[62] T. Dekker, J.M. Rose, Shape shifters: simple asymmetric mixing densities for mixed logit models, IVM Working Paper IVM 11/01 Netherlands (2011).
[63] T. Dekker, The asymmetric triangular distribution for continuous mixed multinomial logit models, in: Transportation Research Board 93rd Annual Meeting. Washington DC, 2014.
[64] K.E. Train, Recreation demand models with taste differences over people, Land Econ. 74 (2) (1998) 230–239.
[65] J.B. Chang, J.L. Lusk, Mixed logit models: accuracy and software choice, J. Appl. Econom. 26 (1) (2011) 167–172.
[66] D. McFadden, Conditional logit analysis of qualitative choice behavior, in: P. Zarembka (Ed.), Frontiers in Econometrics, Academic Press, New York, 1973, pp. 105–142.
[67] J.J. Louviere, D.A. Hensher, J.D. Swait, Stated Choice Methods: Analysis and Applications, Cambridge University Press, Cambridge, UK, 2000.
[68] R. Scarpa, M. Thiene, D.A. Hensher, Preferences for tap water attributes within couples: an exploration of alternative mixed logit parameterizations, Water Resour. Res. 48 (1) (2012).
[69] Y. Gu, R. Norman, R. Viney, Estimating health state utility values from discrete choice Experiments—A QALY space model approach, Health Econ. 23 (9) (2014) 1098–1114.

[70] A.R. Hole, A comparison of approaches to estimating confidence intervals for willingness to pay measures, Health Econ. 16 (8) (2007) 827–840.
[71] I. Olesen, F. Alfnes, M.B. Røra, K. Kolstad, Eliciting consumers' willingness to pay for organic and welfare-labelled salmon in a non-hypothetical choice experiment, Livest. Sci. 127 (2) (2010) 218–226.
[72] I. Krinsky, A.L. Robb, On approximating the statistical properties of elasticities, Rev. Econ. Stat. 64 (4) (1986) 715–719.
[73] I. Krinsky, A.L. Robb, On approximating the statistical properties of elasticities: a correction, Rev. Econ. Stat. 72 (1) (1990) 189–190.