WATER RESOURCES RESEARCH, VOL. 48, W01520, doi:10.1029/2010WR010148, 2012

Preferences for tap water attributes within couples: An exploration of alternative mixed logit parameterizations

Riccardo Scarpa,[1,2] Mara Thiene,[3] and David A. Hensher[4]

Received 21 October 2010; revised 21 November 2011; accepted 3 December 2011; published 19 January 2012.

[1] Preferences for attributes of complex goods may differ substantially among members of households. Some of these goods, such as tap water, are jointly supplied at the household level. This issue of jointness poses a series of theoretical and empirical challenges to economists engaged in empirical nonmarket valuation studies. While a series of results have already been obtained in the literature, the issue of how to empirically measure these differences, and how sensitive the results are to choice of model specification from the same data, is yet to be clearly understood. In this paper we use data from a widely employed form of stated preference survey for multiattribute goods, namely choice experiments. The salient feature of the data collection is that the same choice experiment was applied to both partners of established couples. The analysis focuses on models that simultaneously handle scale as well as preference heterogeneity in marginal rates of substitution (MRS), thereby isolating true differences between members of couples in their MRS, by removing interpersonal variation in scale. The models employed are different parameterizations of the mixed logit model, including the willingness to pay (WTP)-space model and the generalized multinomial logit model. We find that in this sample there is some evidence of significant statistical differences in values between women and men, but these are of small magnitude and only apply to a few attributes.

Citation: Scarpa, R., M. Thiene, and D. A. Hensher (2012), Preferences for tap water attributes within couples: An exploration of alternative mixed logit parameterizations, *Water Resour. Res.*, *48*, W01520, doi:10.1029/2010WR010148.

1. Introduction

[2] Members of households may hold substantially different preferences for attributes of complex goods. Some of these goods, such as tap water, are supplied to all members of the same household. This issue poses a series of theoretical and empirical challenges to economists engaged in empirical nonmarket valuation studies of water supply. In preference studies based on stated choice, these challenges add to the existing ones related to how to parsimoniously model preference heterogeneity. In this paper we present and contrast three parameterizations of the mixed logit model of choice using data obtained from choice experiments on tap water from a sample of 576 respondents (288 couples). The specifications are (1) the conventional mixed logit model with random preference parameters, (2) the generalized mixed logit proposed by *Fiebig et al.* [2010], and (3) the WTP-space mixed logit proposed by *Train and Weeks* [2005]. We discuss the advantages and disadvantages of these models in the context of our study and with reference also to broader issues of modeling taste and WTP variation. We then focus on, and test for, preference diversity between men and women.

[3] While the existing literature provides a series of useful results on individual versus joint choices in couples, the issue of how to empirically measure these differences, and how sensitive the results are to choice of model specification from the same data, is yet to be clearly understood. These issues have important repercussions on the state of practice for surveys of this kind, especially those related to regulated and network industries. Should the household be interviewed or is it sufficient to interview a representative individual? Before delving into the issue of how to represent intrahousehold bargaining for tap water, we believe that the existence of systematic differences between men and women should be ascertained. We set out to investigate this issue in the context of domestic tap water.

[4] In the context of water supply in the UK, for example, since the seminal study by *Willis et al.* [2005] for Yorkshire Water, a number of studies have informed the context of periodical tariff negotiation between water industries and OFWAT. The state of practice of these surveys would be significantly modified if the survey approach, based on the representative agent, is found to be lacking and more reliable preferences estimates are found as a construct of multi-agent interactions resulting in joint decisions. This would signify that once a household is sampled, more than one member of the household should be interviewed to reconstruct the household utility function, perhaps as a function of an intrahousehold bargaining process. From these

[1]Department of Economics, Waikato Management School, University of Waikato, Hamilton 3240, New Zealand.
[2]Centre for the Study of Choice, University of Technology Sydney, NSW, Australia.
[3]Dipartimento Territorio e Sistemi Agroforestali Agripolis, Via dell' Università 16, 35020 Legnaro (PD), Italy.
[4]Institute of Transport and Logistics Studies, The Business School, University of Sydney, NSW 2006, Australia.

0043-1397/12/2010WR010148

estimates one would derive the associated welfare estimates. Evidence from the study of other types of joint decisions, such as choice of holiday destination [*Dosman and Adamowicz*, 2006], beach visitation [*Beharry-Borg et al.*, 2009], automobile purchases [*Hensher et al.*, 2011], and residential location [*Marcucci et al.*, 2011], suggest that joint family or couple decisions may well depart from the representation of preferences obtained from stated choice surveys based on the representative (often randomly sampled) agent of the household.

[5] During the 1990s, a number of discrete choice models have been proposed and have become widely used to address the issue of taste variation across people, and also to overcome the inflexibilities of what had been the workhorse of discrete choice analysis, the multinomial logit choice model. The panel mixed logit model in preference space (henceforth MXL) is perhaps the most frequently used model nowadays when analysts want to address unobserved taste heterogeneity [*Train* 1998, *Hensher et al.* 2005]. It has also been shown that with an appropriate error-component structure inclusive of taste correlation, the MXL approach affords maximum flexibility [*McFadden and Train*, 2000]. The category of mixed logit models is therefore very broad, with many different parameterizations.

[6] In more recent times, specific reparameterizations of the MXL model have been developed to parsimoniously address selected issues, such as the direct derivation of marginal WTP distributions and other marginal rates of substitutions [*Train and Weeks*, 2005], and the parameterization of taste variation with a lognormal mixture of normals [*Fiebig et al.*, 2010]. Such reparameterizations of the basic mixed logit model allows the estimation of the covariance (and hence correlation) structure of taste variation to be achieved in a particularly parsimonious format in terms of additional parameters. Sometimes, especially in classical estimation via maximum simulated likelihood, this is an advantage, given the large number of additional parameters required to estimate all the elements of a Cholesky (or full variance-covariance) matrix.

[7] Our analysis focuses on the effect of using a recent generation of the mixed logit parameterization to address some important issues in the representation of preference variation, and its relationship to the derivation of marginal WTP estimates. The parameterization we employ are the WTP-space model [*Train and Weeks*, 2005] and the generalized multinomial logit model [*Keane* 2006; *Fiebig et al.*, 2009]. All models presented here are estimated on a pooled sample of men and women and then women-specific effects are tested for significance.

[8] The commodity under consideration is the quality of domestic tap water supply at the household level. Specifically, we wish to contribute to the existing literature on water customers' willingness to pay to improve domestic water quality [*Hensher et al.*, 2005; *Willis et al.*, 2005; *Scarpa et al.*, 2007; *Genius et al.*, 2008; *Dupont et al.*, 2009]. Following this literature, we treat domestic water quality as a complex good because its value to the individual or the couple may be decomposed into a number of attributes, yet it is sufficiently simple and certainly familiar to all respondents.

[9] More broadly we also wish to contribute, but only with our empirical evidence, to the debate in the literature on household economics, which has made substantial progress from the initial theoretical contributions [*Becker*, 1973; *Browning and Chiappori*, 1998; *Lampietti*, 1999; *Chiuri*, 2000; *Vermeulen*, 2002], giving rise to a number of empirical applications in various fields, such as marketing [*Arora and Allenby*, 1999; *Adamowicz et al.*, 2005] and transport [*Hensher et al.*, 2007, 2008; *Puckett et al.*, 2007]. With few exceptions [*Quiggin*, 1998; *Smith and Houtven*, 1998; *Bateman and Munroe*, 2005; *Dosman and Adamowicz*, 2006; *Strand*, 2007], empirical investigations of this kind are relatively less developed in environmental economics. Recent studies provide evidence of substantial differences in taste intensities between domestic partners, and make an attempt at reconciling them with observed joint choices using power functions [*Dosman and Adamowicz*, 2006; *Beharry-Borg et al.*, 2009]. The evidence collected so far indicates that for some categories of decisions, the conventional practice of selecting one member of the couple as representative of the tastes of the entire household may be biased when compared with the preference estimates underlying joint deliberation by the same couple. If this were to be confirmed in the case of tap water then the current practice of surveying one representative individual of the household should be reconsidered in favor of surveying more individuals in the household.

[10] The rest of the paper is organized as follows. Section 2 provides details on the survey, data, and experiment design; section 3 describes the methods; and section 4 reports estimation results. Section 5 sets out conclusions and suggestions for future research.

2. Data and Survey

[11] The data are sourced from an explorative survey designed to pave the way for a broader data collection effort that will be the subject of a different study. The survey was carried out in two major cities in the north of Italy, Vicenza in the northeast and Turin in the northwest, as well as a rural area (Arzignano). Vicenza is a typical medium-sized city with middle to high income, while Turin is a larger sized city with middle to low income population. The results reported here are based on interviews of 266 couples, which in total provided 5024 choice responses from 576 respondents. The couples were asked to choose among alternative packages of tap water supply which differed on the basis of four quality attributes relating to drinking water characteristics plus the cost (chlorine odor, chlorine taste, water turbidity, calcium carbonate stains, and cost). Cost was described as an additional amount of money people would pay in the water bill per year, and ranged from a minimum of €2/year to a maximum of €25/year. The levels of this attribute varied across designs. In particular, respondents were asked to choose among water service supply contracts displaying different levels of water supply characteristics. The issue under investigation was of particular relevance because, at the time of the survey, there was an ongoing public debate at various levels of government on the introduction of a new national law representing a turning point in water legislation.

[12] The attributes and the relative levels are reported in Table 1. Respondents were asked to choose between the frequencies of events in which they could SMELL and/or

Table 1. Description of Attributes

Variable Name	Description of Attribute
O_ALWAYS	chlorine odor always
O_MONTH	chlorine odor once a month
O_WEEK	chlorine odor once a week
O_NEVER	chlorine odor never
T_ALWAYS	chlorine taste always
T_MONTH	chlorine taste once a month
T_WEEK	chlorine taste once a week
T_NEVER	chlorine taste never
NO_TURB	no turbidity from fine air bubbles
MILD_TURB	mild turbidity from fine air bubbles
MED_TURB	medium turbidity from fine air bubbles
EXTR_TURB	extreme turbidity from fine air bubbles
STAIN	presence of calcium carbonate staining in pipes

TASTE chlorine (once a day, once a week, once a month, never, or always). TURBIDITY due to fine air bubbles was also considered; with levels including its absence and its presence in a mild, medium, and extreme form. Due to the hardness of water in this area, calcium carbonate STAINING in pipes is a big issue, and the effect of presence/absence of staining was also investigated.

[13] Respondents faced four alternatives in each choice set, where one alternative was always the status quo and involved no additional cost. This gave respondents the option of defaulting on the choices when the attributes' levels in the proposed alternatives were not sufficiently desirable to them. An example of a choice set is reported in Table 2.

[14] In the survey, respondents were asked to express their preferences toward the main types of water supply, that is, tap, bottle, and home filtered water; the reasons behind their choice; and their level of satisfaction with tap water. The specific preference elicitation question posed at each choice task to the individual respondent was: "Which of the following four alternatives would you choose?" This question was aimed at eliciting individual preferences and not household preferences. On the other hand, when interviewed jointly as a couple, the questions referred to their joint preferences. Some socioeconomic data, including household income, were also collected.

[15] Experts employed by local utilities supplying Integrated Water Services were contacted and interviewed in order to finalize the design of the survey. They provided specific and technical information which was extremely valuable in the selection of the attribute levels. This information was integrated with suggestions provided by public institutions involved in the management of water service. The combined information was then used to conduct repeated focus groups, the results from which were then used to design the choice experiments. The complete questionnaire was then tested in the field by means of a pilot survey.

2.1. Sequential Experimental Design

[16] There is a growing literature describing efficiency criteria that can be maximized to ensure high accuracy in estimates from stated choice data [*Sandor and Wedel*, 2001; *Ferrini and Scarpa*, 2007; *Rose and Bleimer*, 2009]. This becomes particularly relevant when considering that experimental design can significantly impact the efficiency of WTP estimates, which play a central role in nonmarket valuation [*Lusk and Norwood*, 2005; *Scarpa and Rose*, 2008; *Vermeulen et al.*, 2011].

[17] Efficient designs for logit model specifications depend on accurate priors on the population parameters. During the course of the survey, as more responses are collected, gradually more information becomes available on such parameters. Hence it is possible to increase the efficiency of the final estimates by adopting a sequential experimental design, rather than relying on one single design for the entire survey [*Kanninen*, 1993; *Scarpa et al.*, 2007; *Kerr and Sharp*, 2010]. Given the need for accurate WTP estimates for policy inference, our attention was placed on *c* efficiency, hence a Bayesian criterion, namely the WTP_b*-efficient* criterion was adopted to select the fraction of the full factorial to be used as a design in the sequence of subsamples. This criterion consists of the minimization of the variance of some nonlinear functions of the utility coefficients, namely the sum of the variances of the marginal willingness to pay estimates. Taking into account that different attributes can be described in different units, *Scarpa and Rose* [2008] point out that the minimization process of the sum of variances of marginal WTPs with an uneven unit of measurement may result in an unsatisfactory outcome. To overcome such a limitation, they suggest the adoption of a criterion that maximizes the minimum *t* value for the marginal WTP. This choice places more emphasis on the attribute whose WTP was estimated with least accuracy, as measured by the *t* value. We note in passing that Bayesian WTP efficiency has also been found to provide designs with higher robustness to outliers and are less prone to producing extreme WTP estimates [*Vermeulen et al.*, 2011].

[18] The overall survey design was articulated in subsequent phases, as gradually additional information was sequentially collected in six waves. Each sample wave used a different WTP_b*-efficient* design developed using Bayesian priors (as indicated by the subscript "b"), derived by combining the information collected in all previous waves. The initial prior information was gathered from the pretest and the pilot survey; the first wave of interviews then informed in turn the design of the following waves. At the end of waves 1–6, the data was coded and basic multinomial logit models were estimated so as to provide priors for the

Table 2. Illustrative Choice Set

Which of the Following Alternative Would You Choose?	A	B	C	D
Chlorine odor	Always	1 day per week	1 day per month	None
Chlorine taste	Always	1 day per week	Never	
Turbidity	Absent	Medium	Extreme	
Calcium carbonate staining	No	Yes	Yes	
Additional WTP in the bill per year	18€	5€	6€	
Mark choice				

efficient design of the subsequent sample wave. Each respondent answered either eight or nine choice tasks.

2.2. Survey and Sample

[19] Sampling was conducted among households in two cities—Vicenza and Turin—as well as a rural area (Arzignano, close to Vicenza). The main objective of the study was to provide a preliminary investigation of the preferences of individuals, and to ascertain if, in the aggregate, men and women display similar tastes. We must note that the sample cannot be taken as representative of the residential population because, although respondents were randomly selected into the sample, it was not sufficiently broad in geographical terms. We hence acknowledge the consequent limitations of the study in terms of its policy implications. However, we cannot find any obvious link between the selection mechanism into the sample and the results we report, which focuses on differences across specifications and WTP distributions between genders.

[20] The survey involved several stages. The first stage was aimed at selecting households that could be considered as "couples," that is with a stable relationship. During the second stage, they were asked whether they would be willing to participate in the survey. They were contacted by telephone. Once both partners agreed to participate, the interviewer would make an appointment to visit the couple. At the couple's home they were debriefed jointly and given the stated preference tasks which were executed independently.

3. Method

[21] Our starting point in the analysis of the data is the panel MXL model of multinomial choice proposed by *Train* [1998], where the utility of alternatives is represented by

$$U_{njt} = \lambda V_{njt} + \varepsilon_{njt} = \lambda \boldsymbol{\alpha}'_n \mathbf{x}_{njt} + \varepsilon_{njt} = \boldsymbol{\gamma}'_n \mathbf{x}_{njt} + \varepsilon_{njt},\ \lambda > 0 \quad (1)$$

and n denotes the respondent, j is the alternative, t is the sequence in the panel, λ is the scale of the i.i.d. Gumbel error denoted by ε_{njt}. V_{njt} is the indirect utility and $\boldsymbol{\alpha}_n$ is one realization of the vector of random taste intensities for respondent n. As shown by *McFadden and Train* [2000] under the assumption that $\boldsymbol{\alpha}$ can take *any* multivariate distribution, inclusive of all possible correlation structures between its elements, (1) can approximate any random utility model.

[22] While the implications of various distributional assumptions of the random variable $\boldsymbol{\alpha}$ have been the subject of much discussion in the literature [see *Daly et al.*, 2011 for a review], and similarly for structural dependencies of λ [e.g., $\lambda = \exp(\boldsymbol{\theta}'\mathbf{z})$, where the vector $\mathbf{z}$ may contain the number of alternatives, measures of choice complexity, order in the sequence, entropy measures, etc., and some of the elements of $\boldsymbol{\theta}$ may be assumed to be random], less attention has been devoted to modeling the interpersonal unobserved variation of λ, which captures the joint correlation across all elements of the random vector $\boldsymbol{\alpha}$. It is clear that due to the positive nature of λ, and to the multiplicative relationship between $\boldsymbol{\alpha}$ and λ, the random vector $\boldsymbol{\gamma}$ will be correlated with λ when the latter is random. In fact, a random λ will absorb much of the jointness of variation across $\boldsymbol{\gamma}$, and hence the existing correlation. Since some form of correlation structure across random marginal utilities (or part-worths) is to be expected in most problems, the analyst can choose to either model the variation of $\boldsymbol{\gamma}$, or to model the variation of $\boldsymbol{\alpha}$ and λ separately, thereby implying a given correlation in $\boldsymbol{\gamma}$. While the first option is the most flexible, the second option is attractive from both the pedagogical viewpoint and for its parsimony in terms of the number of parameters that need to be estimated. These advantages are important, especially when estimation proceeds via simulated maximum likelihood rather than by Monte Carlo Markov chains. It is clearly the same feature of randomness that underlies the real data that is to be modeled, i.e., correlation across part-worths. But it is achieved with different parameterizations. In short, total unobserved taste variation can sometimes be conceptually separated into two components, with some joint and perfectly correlated variation across individuals to be ascribed to scale and the remainder—residual and imperfectly correlated—to be ascribed to taste [*Swait and Louviere*, 1993; *Louviere and Eagle*, 2006]. It is clear though that the two cannot be separated since they exist jointly. Their separation in estimation is only an artificial construct. For the purpose of derivation of population WTPs, we note how the preference-space MXL with independent normal coefficients will produce a random structure of WTPs which is correlated by means of the shared random cost coefficient in the denominator [*Train and Weeks*, 2005]. With a fixed cost coefficient and independent elements in the random vector $\boldsymbol{\gamma}$, this correlation will be zero.

[23] With this perspective in mind, at one end of the spectrum of model forms, with perfectly correlated part-worths, we have the random scale multinomial logit model (or SML) [*Breffle and Morey*, 2000], which only assumes the presence of a varying scale parameter, conditional on which respondents are all "preference clones" up to a idiosyncratic i.i.d. Gumbel error. The SML utility structure is

$$U_{njt} = \lambda_n V_{jt} + \varepsilon_{njt} = \lambda_n \boldsymbol{\alpha}' \mathbf{x}_{njt} + \varepsilon_{njt} = \boldsymbol{\gamma}'_n \mathbf{x}_{njt} + \varepsilon_{njt},\ \lambda_n > 0. \quad (2)$$

Unlike in equation (1), the SML model is estimated assuming a random rather than a constant λ. Typically λ is log-normally distributed, and in that case the assumed population parameters for the attributes are also distributed log-normally and shifted along the orthant by the nonrandom location parameter α_k, where k refers to the attribute. However, other distributions for scale can be handled by the SML, as long as they are defined over the positive orthant. A drawback of this model is that of imposing perfect correlation across the resulting part-worths, and for this reason we do not estimate this model here.

[24] At the other end of the spectrum we have the MXL with zero off-diagonal elements in the variance-covariance matrix. The utility structure is as in equation (2), and in this case each random coefficient can take a different distribution to accommodate prior beliefs or theoretical expectations. In most real cases though, taste for attributes do not vary independently and are correlated. Ignoring such correlation has been shown to bias estimates.

[25] It has recently been suggested that it is sometimes useful to separate scale randomness from randomness of marginal utility coefficients. This can be achieved under

relatively restrictive distributional assumptions. Two models emerge in this literature: the generalized multinomial logit model (G-MNL) proposed by *Keane* [2006] and the WTP-space model proposed by *Train and Weeks* [2005]. Both models have merits and drawbacks. The former is attractive because it nests all other models with lognormal mixtures of normally distributed utility coefficients. Since the normal distribution is often used in practical work, this parameterization can be useful. However, in many instances normality is a poor approximation of what is expected in reality. The WTP-space model does not suffer from the same restriction, thereby accommodating a wider range of distributional assumptions on marginal WTPs, and it directly obtains population estimates for these, which might be convenient in nonmarket valuation studies. However, as a drawback it restricts the cost coefficient to be perfectly correlated with scale. A specific form of the WTP-space model with normal marginal WTPs can be derived from the G-MNL specification by imposing some restrictions. When other distributional assumptions for marginal WTPs are necessary such restrictions on the G-MNL will not achieve the WTP-space model.

[26] Consider the G-MNL model utility specification as from equation (7) by *Fiebig et al.* [2009]:

$$U_{njt} = \boldsymbol{\alpha}_n' \mathbf{x}_{njt} + \varepsilon_{njt} = [\lambda_n \boldsymbol{\beta} + \gamma \boldsymbol{\eta}_n + (1-\gamma)\lambda_n \boldsymbol{\eta}_n]\mathbf{x}_{njt} + \varepsilon_{njt}, \quad (3)$$

where, using the same notation as in the seminal paper, λ_n is lognormal and η_n is a zero-mean normal vector.

[27] By setting $\theta = 0$, we obtain

$$U_{njt} = [\lambda_n \boldsymbol{\beta} + \lambda_n \boldsymbol{\eta}_n]\mathbf{x}_{njt} + \varepsilon_{njt} = [\beta + \boldsymbol{\eta}_n]\lambda_n \mathbf{x}_{njt} + \varepsilon_{njt} = \tilde{\boldsymbol{\beta}}_n \lambda_n \mathbf{x}_{njt} + \varepsilon_{njt}. \quad (4)$$

With the further restriction of the cost coefficient to be equal to -1, one obtains the WTP-space model with normal WTPs [see *Train and Weeks*, 2005 for details]:

$$U_{njt} = \lambda_n^* (\boldsymbol{\omega}_n' \mathbf{x}_{njt} - p_{jt}) + \varepsilon_{njt}, \quad (5)$$

where $\mathbf{x}$ now includes only the nonprice attributes, p_{jt} denotes price, $\boldsymbol{\omega}_n$ is the realization for respondent n of the random vector of marginal WTPs for each nonprice attribute, and $\lambda_n^* = \lambda_n \delta_n$, where δ_n is the realization for respondent n of the price coefficient. A specific covariance structure can be estimated to account for correlation across the random normal $\boldsymbol{\omega}_n$.

[28] As mentioned earlier, the G-MNL is a generalization of a special category of MXL, that in which the distribution of utility coefficients is a lognormal mixture of normals. Note that even when the idiosyncratic components in the vector $\boldsymbol{\eta}_n$ are independent, the vector $\tilde{\boldsymbol{\beta}}_n \lambda_n = \boldsymbol{\alpha}_n$ will have correlated elements through the joint variation of the lognormal λ_n, while $\tilde{\boldsymbol{\beta}}_n$ are normally distributed. Other distributional assumptions of $\boldsymbol{\alpha}_n$ which can be readily implemented in the MXL are not compatible with the G-MNL specification. So, despite the claims made in *Fiebig et al.* [2010], at close inspection the G-MNL appears to be not as general as the paper purports. It is nevertheless a useful simplification in contexts in which the assumed distribution is coherent with the data at hand and one is interested in introducing correlation across the elements of $\boldsymbol{\alpha}_n$ in a parsimonious manner, as this can be achieved by the estimation of the distributional moments of one single random parameter λ_n.

[29] In estimation, to ensure the condition $\lambda_n > 0$, lognormality is assumed by setting $\lambda_n = \exp(\overline{\lambda} + \tau\eta_{0n})$ with $\eta_{0n} \sim N(0,1)$. Because of its multiplicative role, the mean value of λ_n is set to one for normalization [*Fiebig et al.*, 2010]. Since $E\lambda_n = \exp\left(\overline{\lambda} + \frac{\tau^2}{2}\right)$, and τ is to be estimated, during the estimation by simulation, a mean value of 1 is achieved by centering the lognormal distribution by setting $\overline{\lambda} = -\ln\left(\frac{1}{R}\sum_{r=1}^{R} \exp(\tau\eta_0^r)\right)$, where τ is the estimated standard deviation from the data, and η_0^r is a draw from standard normal truncated at $[-2, 2]$, so as to avoid extreme values [*Fiebig et al.*, 2010].

[30] Willingness to pay and welfare estimates in general can be derived by using closed-form solutions based on utility coefficients from preference space specifications [*Train*, 2003]. Such closed forms are characterized by the cost coefficient in the denominator of a ratio. When this coefficient is assumed to be random and defined over an interval including zero, a draw with values very close to zero will make the ratio "explode" into a very large number. A random denominator also produces problems in deriving closed form means and variances of the ratio of random variables. The presence of such extreme values complicates the computation of the mean and of the variance of the WTP distributions in simulations used to approximate them (the so-called inexistence of distribution moments "trap"). The existence of the central moments of the implied WTP distributions is hence dependent on the distributional features of the cost coefficient (the marginal utility of money). *Daly et al.* [2011] provide a useful and clear discussion of the conditions under which these moments exist, with various examples. In principle though, any continuous density with nonzero values around the origin will imply nonexistence of central moments. However, a characterization of the population distributions of WTPs can still be achieved if one can be satisfied with a quantile-based representation, as distribution quantiles are robust to outliers, unlike the central moments of a distribution.

[31] Analysts interested in WTP distributions with finite moments might find it more attractive to use specifications in WTP space than to use G-MNL specifications with random normal cost coefficients. This is because in the G-MNL case the distribution spans the neighborhood of zero and finite means and variance of the implied WTP distributions do not exist. Whether preference space or WTP-space specifications fit the data better is an empirical issue, with evidence in favor of both cases [*Scarpa et al.*, 2008; *Thiene and Scarpa*, 2009; *Balcombe et al.* 2009, etc.]. Alternatively, one can fix the cost coefficient, but this is less than ideal because it implies constant marginal utility of money across the population, which is likely to be an implausible assumption in many practical cases. G-MNL and any MXL model with all normal coefficients are poorly suited to model choices between alternatives when attribute levels are expected to improve (worsen) on some status quo condition. The expectation is that these coefficients should be positive (negative) while the normal distribution

typically spans the real line. In our study, attribute values are ordinal and not only do we have clear expectations on the sign of the coefficients, but we also expect some relative ordering in their magnitude. Clearly the degree of severity of this source of misspecification very much depends on whether the estimated standard deviation coefficient for the normal random parameter is sufficiently large to allocate a large portion of the implied taste distribution to the orthant with the unexpected sign.

4. Results

[32] Table 3 reports the model estimates for three models, all of which are estimated from the pooled sample of men and women using dummy coded attribute levels. For each attribute the worst condition is omitted and used as the baseline, so that attribute coefficients are all expected to be positive. All models were estimated using at least 500 pseudo-random draws of the Latin-hypercube sampling type [*Hess et al.*, 2006]. The columns on the left report estimates for a MXL model with uncorrelated normal random coefficients and a fixed cost coefficient, which imply independent normal marginal WTP distributions. We also estimated models with a lognormal price coefficient (MXL full correlation $\ln L$ −2,453, without correlation $\ln L$ −2,734), which we do not report here as they imply similar value distributions, and afford only a small improvement in fit with regards to the models presented here. What is noticeable from the MXL model in Table 3 is the lack of evidence in support of taste variation for some of the attribute levels which have insignificant standard deviations (e.g., for odor once a week and once a month, and for no turbidity).

[33] The G-MNL model estimates are in the central columns of Table 3. This specification is a direct extension of the MXL with normal coefficients, but it allows for correlation of taste and, by implication, of marginal WTPs by including λ as an additional parameter, which in this context we call "scale," following *Fiebig et al.* [2010]. In this parameterization the standard deviations are more frequently significant. Note that to estimate this model the cost variable was divided by 10 so as to avoid overflow problems in computation.

[34] The two columns on the right report estimates of a model in WTP space (estimated scaling cost by 1/100) with lognormal λ_n^* and independent normal marginal WTPs [*Train and Weeks*, 2005]. These model estimates show significant heterogeneity for all the random coefficients. We note that this model outperforms, in terms of fit, the MXL with normal random coefficients, and it is more than halfway closer to the fit of the G-MNL than the MXL model.

[35] In terms of estimated means across models, these have the expected signs and expected relative magnitudes in the conventional preference space MXL and in the WTP-space reparameterization. The expected relative magnitude, instead, is lost in the G-MNL estimates, where the coefficient for the mean of experiencing *odor once a week* is larger than that for the lower frequency of *once per month*. This also happens for the *taste* levels, but not for

Table 3. Estimates for Uncorrelated Models (|t-values| in Parentheses), 576 Respondents, 5024 Choices

Parameters		MXL		G-MNL		WTP-S (in €100)	
Odor once a week	μ	1.18	(7.2)	5.39	(4.6)	0.10	(46.0)
	σ	0.24	(0.9)	0.37	(0.8)	0.03	(61.8)
Odor once a month	μ	1.64	(10.6)	3.96	(4.2)	0.14	(53.9)
	σ	0.14	(0.6)	0.57	(1.9)	0.05	(47.9)
Odor never	μ	1.88	(11.7)	6.55	(4.8)	0.16	(61.7)
	σ	0.84	(4.9)	1.23	(4.2)	0.06	(48.3)
Taste once a week	μ	0.24	(1.2)	2.29	(4.1)	0.04	(12.2)
	σ	0.80	(2.9)	0.84	(2.5)	0.02	(13.4)
Taste once a month	μ	1.20	(6.9)	1.62	(3.2)	0.06	(28.4)
	σ	0.45	(1.8)	0.02	(0.1)	0.08	(99.9)
Taste never	μ	2.06	(13.9)	5.55	(5.2)	0.14	(52.6)
	σ	0.04	(0.1)	0.29	(0.9)	0.01	(7.0)
Medium turbidity	μ	0.06	(0.3)	1.76	(2.3)	0.07	(22.7)
	σ	0.94	(3.9)	0.70	(2.7)	0.09	(75.7)
Mild turbidity	μ	0.54	(2.1)	4.11	(3.9)	0.11	(35.3)
	σ	1.42	(7.3)	0.82	(2.8)	0.09	(55.0)
No turbidity	μ	2.87	(17.3)	10.74	(4.7)	0.32	(96.7)
	σ	0.71	(4.6)	0.42	(1.7)	0.01	(22.6)
No Stain	μ	1.63	(15.7)	7.22	(4.8)	0.13	(55.4)
	σ	0.77	(5.5)	1.14	(5.0)	0.08	(22.5)
Cost	μ	−0.12	(18.8)	−4.91	(4.8)		
	σ			2.00	(6.1)		
ln(λ)	μ					−3.96	(16.2)
	σ					2.20	(7.7)
Status quo	μ	6.44	(26.3)	13.37	(4.9)	0.56	(69.3)
		1.98	(17.1)	2.45	(6.5)	0.00	(2.0)
Variance parameter in scale	τ			1.24	(9.0)		
Weighting parameter gamma	γ			0.98	(1.4)		
ln-Likelihood		−2761.87		−2460.98		−2604.03	
Pseudo-R^2		0.60		0.65		0.63	
Bayesian information criteria (BIC)		5669.93		5087.22		5360.60	
Akaiki information criteria (AIC)		5569.74		4973.96		5256.06	
Akaiki information criteria 3 (AIC3)		5592.74		4999.96		5280.06	
		576 respondents, 5024 choices					

Table 4. Models With Correlated Coefficients (|t-values| in Parentheses), 576 Respondents, 5024 Choices

Parameters		MXL		WTP-S (in €100)	
Odor once a week	μ	1.96	(12.2)	0.25	(35.0)
	σ			0.07	(75.8)
Odor once a month	μ	2.46	(24.7)	0.32	(45.8)
	σ			0.01	(9.3)
Odor never	μ	2.86	(27.5)	0.33	(46.4)
	σ			0.06	(82.0)
Taste once a week	μ	0.68	(4.6)	0.04	(9.8)
	σ				
Taste once a month	μ	1.06	(7.4)	0.05	(29.1)
	σ			0.05	(50.7)
Taste never	μ	1.98	(15.8)	0.15	(94.0)
	σ			0.00	(2.5)
Medium turbidity	μ	0.64	(3.8)	−0.01	(0.9)
	σ			0.10	(30.7)
Mild turbidity	μ	1.25	(7.5)	−0.03	(2.9)
	σ			0.16	(50.7)
No turbidity	μ	4.89	(22.1)	0.30	(45.0)
	σ			0.01	(30.7)
No stain	μ	3.02	(11.2)	0.14	(86.1)
	σ	1.97	(9.8)	0.07	(51.9)
Cost	μ	−0.12	(15.4)		
ln(λ)	μ			−4.77	(17.9)
	σ			2.82	(10.7)
Status quo	μ	10.19	(27.7)	0.71	(0.0)
	σ	3.50	(0.0)	0.00	(2.3)
Cholesky (stain, SQ)		−3.79	(17.5)		
Cholesky (odor week, no stain)		−0.32	(2.4)	0.12	(27.6)
Cholesky (odor week, SQ)		−0.56	(4.4)	0.01	(8.4)
Cholesky (odor week, odor month)				−0.01	(6.2)
Cholesky (odor week, odor never)		−0.90	(6.4)		
Cholesky (odor week, taste week)				0.03	(11.2)
Cholesky (odor week, taste month)				0.02	(15.5)
Cholesky (odor week, taste never)				0.05	(38.6)
Cholesky (odor week, medium turbidity)				0.05	(35.7)
Cholesky (odor week, mild turbidity)				0.06	(30.7)
Cholesky (odor week, no turbidity)				0.02	(20.6)
Cholesky (odor month, no stain)				0.11	(26.0)
Cholesky (odor month, SQ)		−0.85	(12.3)	−0.01	(7.8)
Cholesky (odor month, odor never)				0.00	(3.3)
Cholesky (odor month, taste week)				−0.02	(11.4)
Cholesky (odor month, taste month)				0.01	(11.4)
Cholesky (odor month, taste never)				0.01	(9.3)
Cholesky (odor month, medium turbidity)				−0.02	(15.9)
Cholesky (odor month, no turbidity)		−0.47	(5.4)	0.00	(5.2)
Cholesky (odor never, no stain)				0.14	(34.3)
Cholesky (odor never, SQ)		−0.92	(0.0)		
Cholesky (odor never, taste week)				−0.03	(17.0)
Cholesky (odor never, taste month)				−0.07	(54.7)
Cholesky (odor never, taste never)				−0.06	(46.9)
Cholesky (odor never, mild turbidity)				−0.01	(4.2)
Cholesky (odor never, no turbidity)				0.07	(56.5)
Cholesky (taste week, SQ)				0.02	(11.2)
Cholesky (taste week, taste month)				−0.05	(49.2)
Cholesky (taste week, taste never)				−0.03	(53.0)
Cholesky (taste week, medium turbidity)				−0.04	(20.9)
Cholesky (taste week, mild turbidity)				−0.03	(17.7)
Cholesky (taste month, no stain)				0.03	(18.8)
Cholesky (taste month, SQ)				0.01	(4.6)
Cholesky (taste month, taste never)				−0.02	(24.4)
Cholesky (taste month, no turbidity)				−0.01	(8.8)
Cholesky (taste never, no stain)				0.01	(11.8)
Cholesky (taste never, SQ)				0.00	(3.1)
Cholesky (taste never, medium turbidity)				−0.01	(9.3)
Cholesky (taste never, no turbidity)				0.00	(4.4)
Cholesky (medium turbidity, no turbidity)				0.02	(54.0)
Cholesky (mild turbidity, no stain)				0.00	(1.6)
Cholesky (mild turbidity, no turbidity)				−0.02	(67.2)
Cholesky (no turbidity, no Stain)		−1.75	(29.3)	0.01	(3.7)
Cholesky (no turbidity, SQ)		−1.35	(7.3)		
ln-Likelihood		−2546.41		−2527.25	

Table 4. (continued)

Parameters	MXL	WTP-S (in €100)
Pseudo-R^2	0.63	0.64
Bayesian information criteria (BIC)	5239.01	5461.29
Akaiki information criteria (AIC)	5138.82	5182.50
Akaiki information criteria 3 (AIC3)	5161.82	5246.50
576 respondents, 5024 choices		

those of *turbidity*. Despite this, the best fitting model is the G-MNL; the large value estimate of γ^*, which refers to γ^* in the expression $\gamma = [1 + \exp(-\gamma^*)]^{-1}$, suggests that γ in equation (3) is about 0.73. We are hence in the presence of a "true" G-MNL model in which the independent normal idiosyncratic departures from the means are mostly unscaled. The mixture of lognormal and normals is in part (27%) unscaled by the scale parameter and adds directly to the mean, and in the remaining part (73%) added to the mean as scaled. In a simulation of the implied distributions for the G-MNL, we find that the variation of the scale parameter dominates over the relatively small variation of the idiosyncratic random normal effects, so that the vector of composite random tastes $\tilde{\beta}_n$ has extremely high positive correlations ($>$0.97). We cannot comment on the mean and variance of the distributions of marginal WTPs of this model as they do not exist. However, the percentiles of the simulated distributions are all well-behaved (available from the authors). Unfortunately though, like the means, the distributions do not respect the expected relative order of improvements, and because of these discouraging results, we focus the rest of the analysis on extending the MXL and the WTP-space models to include correlation between random coefficients. Similar to the finding of *Greene and Hensher* [2010], we find that accounting for "scale" variation does not make too much of difference, and it somewhat worsens some of the properties of the implied distributions of WTP found in both the MXL and WTP-space models.

[36] Table 4 reports two selected models from a specification search in which we started from the MXL with fixed cost coefficient (to avoid the inexistence of moments "trap" for the WTP distributions) and the WTP-space model in the previous table, and added initially a full Cholesky matrix (MXL ln$L$ −2,491 and WTP-S ln$L$ −2,541). To add more structure, we proceeded by gradually eliminating all the elements of such a matrix that emerged to be less significant than a t value of 1.8. In Table 5 σ represents the diagonal elements of the Cholesky matrix, and for the MXL model we also find nine terms in the Cholesky matrix to be significant, but many diagonal terms of the Cholesky matrix were dropped in the final model. For the WTP space model—instead—only three Cholesky terms are found to be not significant; namely the marginal WTP-S of no-stain and the status-quo ASC, those of *odor never* and *once a week* and *odor never* and the *status quo ASC*. We note though that due to parameter proliferations the AIC and BIC favor the specification without correlation for the WTP-space model. Both models in Table 4 are concordant in suggesting that reducing the frequency of chlorine odor occurrence is the most valuable action to this sample. This is followed by the reduction of turbidity by fine air bubbles, which is twice as valuable as both the absence of calcium carbonate stains and the reduction to zero frequency of chlorine taste. Note though that the value of reduction—as opposed to elimination—of turbidity is basically zero. On the other hand, the reductions of frequency of taste of chlorine from persistent to "once per month" or "once a week" are basically worth the same. The latter are worth approximately one third of the former, according to the WTP-S model.

Table 5. Correlations Between Random Parameters From MXL Model

	Odor Never	No Turbidity	No Stain
Odor never	1.00		
No turbidity	0.00	1.00	
No stain	0.12	0.00	1.00
SQ	0.10	0.15	−0.34

[37] Because the values of the Cholesky matrix are not very meaningful per-se, we report the implied correlation matrices for both models in Tables 5 and 6. While the correlation values for the MXL model are small, those between the implied WTP distributions from the WTP space are sometimes high. WTPs for *no turbidity* have a negative correlation of −0.87 with *odor never*, suggesting that those who have high WTP for one have relatively low WTP for the other; while positive correlations of 0.7 are found between *taste of chlorine once a month* and *never* experiencing its *odor*, and the same values is found for WTP for *no stain* and *no turbidity*.

[38] Our results indicate that there is no substantial difference in WTP estimates between the three models, but the WTP-S appears a good compromise between data fit, expected relative magnitudes, correlation structure, and parsimony. Allowing the WTP model to account for women-attribute interactions (Table 7) requires the additional inclusions of ten parameters, but improves the log likelihood from a value of −2,604.03 to −2,591.51, which gives a p value of less than 0.01 on the test of joint significance of the interaction coefficients, and clearly provides a significant improvement. We therefore conclude that there is statistical evidence of gender differences in WTP space from this sample. However, the economic magnitudes of these differences are small. In particular, women are willing to pay €2 less than men for lowering the frequency to once a week, and the same effect is found for once a month. The other effects are all insignificant, with the exception of no stain, for which women are willing to pay €2 more than men.

[39] Overall, it would appear that preferences for tap water attributes are quite homogeneous across genders, with the exception of water turbidity from air bubbles and absence of calcium carbonate deposit (stains). This heterogeneity is not surprising, as both attributes have levels that are quite differently distributed in the area from which we sampled. So, some variation is to be expected. In terms of

Table 6. Correlation Between Random WTPs From Model in WTP Space

	Odor Once a Week	Odor Once a Month	Odor Never	Taste Once a Week	Taste Once a Month	Taste Never	Medium Turbidity	Mild Turbidity	No Turbidity	No Stain	Cost
Odor once a week	1.00										
Odor once a month	0.46	1.00									
Odor never	0.00	0.04	1.00								
Taste once a week	−0.63	−0.63	0.66	1.00							
Taste once a month	−0.20	0.02	0.70	0.55	1.00						
Taste never	−0.57	−0.17	0.70	0.79	0.89	1.00					
Medium turbidity	−0.43	−0.36	−0.01	0.34	0.22	0.33	1.00				
Mild turbidity	−0.36	−0.17	0.05	0.27	0.21	0.31	0.22	1.00			
No turbidity	−0.26	−0.07	−0.87	−0.45	−0.51	−0.44	−0.10	0.30	1.00		
No stain	−0.51	0.19	−0.60	−0.29	−0.33	−0.13	0.13	0.16	0.70	1.00	
Cost	−0.38	−0.43	−0.01	0.35	−0.50	−0.16	−0.04	−0.02	0.06	0.10	1.00

"bang for the buck" the most beneficial improvement at the margin is *absence of odor* with €33/year followed by the *absence of turbidity*, with about €28/year.

5. Conclusions

[40] Using a purpose-designed choice experiment conducted among members of couples, in this study we explored differences in estimated taste for tap water quality between males and females. We explored the use of different parameterizations of the basic mixed logit model and found that for our purpose the WTP-space specifications provide a parsimonious solution that fits the data well and give plausible results while addressing both Gumbel scale and taste variation. The basic mixed logit with a nonrandom cost coefficient gives the worst fit, but provides similar WTP estimates. The G-MNL gives the best fit, but the implied WTP distributions have an unexpected relative magnitude and are excessively correlated. Using the WTP-space model we test, via maximum simulated likelihood estimation, for the presence of differences in means of marginal WTP distributions for tap water attributes for men and women, and we reject the null of no difference. So, marginal WTPs are significantly different between men and women. In terms of economic magnitudes though, these differences are small and individually significant only for a subset of attributes, which include low and medium frequencies of smell of chlorine and absence of staining from calcium carbonate.

Table 7. Estimates of Model With Women Effect (|t-values| in Parentheses)

Parameters		WTP-S (in €100) Beta		Women Effect	
Odor once a week	μ	0.12	(12.8)	−0.02	(2.5)
	σ	0.02	(7.9)		
Odor once a month	μ	0.16	(15.2)	−0.02	(2.8)
	σ	0.01	(7.3)		
Odor never	μ	0.16	(15.6)	0.00	(0.3)
	σ	0.06	(30.9)		
Taste once a week	μ	0.05	(10.7)	−0.00	(0.5)
	σ	0.03	(13.9)		
Taste once a month	μ	0.05	(27.9)	0.00	(0.2)
	σ	0.08	(54.7)		
Taste never	μ	0.15	(69.4)	−0.01	(1.7)
	σ	0.00	(2.0)		
Medium turbidity	μ	0.01	(1.0)	0.00	(0.1)
	σ	0.11	(43.5)		
Mild turbidity	μ	0.05	(8.3)	0.03	(1.1)
	σ	0.10	(45.1)		
No turbidity	μ	0.28	(32.0)	0.01	(0.4)
	σ	0.00	(2.2)		
No stain	μ	0.12	(51.3)	0.02	(2.0)
	σ	0.09	(53.1)		
ln(λ)	μ	−4.05	(5.2)		
	σ	2.14	(2.7)		
Status quo	μ	0.52	(28.7)		
		0.02	(21.2)		
ln-Likelihood			−2591.51		
Pseudo-R^2			0.63		
Bayesian information criteria (BIC)			5399.13		
Akaiki information criteria (AIC)			5251.02		
Akaiki information criteria 3 (AIC3)			5285.02		
576 respondents, 5024 choices					

[41] Some more specific conclusions can be offered with respect to the practice of using one member of the couple as a representative of the couple's preference to estimate random utility models from survey data in water utilities studies. While this finding is derived with a relatively limited sample, and it is therefore in need of further confirmation from other applications, it would appear that the current practice of administering the survey to one member of the couple, which is often the woman, should not substantially affect the economic dimension of willingness to pay estimates for tap water for most attributes. It appears that the economic magnitude of a couple's preferences over tap water attributes is well represented by sampling one member of the couple, regardless of the sex. Further research work on preferences for tap water should also investigate the role of other members of the family, such as dependents, to provide a clear picture of how their preferences interact with those of the other members of the household. On the econometric side, given the observed differences in model performance, further research might be called for to evaluate these models, perhaps by using Monte Carlo studies, rather than an application to a real data set. This would move us somewhere closer to answering the question of under what conditions one can expect significant differences between these alternative mixed logit specifications.

[42] **Acknowledgments.** We are thankful for comments provided by three anonymous reviewers, which helped us clarify our thoughts on various issues and, as a consequence, substantially improved the paper. The usual disclaimer applies.

References

Adamowicz, W., M. Hanemann, J. Swait, R. Johnson, D. Layton, M. Regenwetter, T. Reimer, and R. Sorkin (2005), Decision strategy and structure in households: A groups perspective, *Marketing Lett.*, *16*, 387–399.

Arora, N., and G. Allenby (1999), Measuring the influence of individual preference structures in group decision making, *J. Marketing Res.*, *36*(4), 476–487.

Balcombe, K., A. Chalak, and I. M. Fraser (2009), Model selection for the mixed logit with Bayesian estimation, *J. Environ. Econ. Manage.*, *57*, 226–237.

Bateman, I., and A. Munro (2005), An experiment on risky choice amongst households, *Econ. J.*, *115*, 176–189.

Becker, G. (1973), A theory of marriage: Part I, *J. Polit. Econ.*, *81*, 813–846.

Beharry-Borg, N., D. A. Hensher, and R. Scarpa (2009), An analytical framework for joint vs separate decisions by couples in choice experiments: The case of coastal water quality in Tobago, *Environ. Resour. Econ.*, *43*, 95–117.

Breffle, W. S., and E. R. Morey (2000), Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of Atlantic salmon fishing, *Marine Resour. Econ.*, *15*, 1–20.

Browning, M., and P. A. Chiappori (1998), Efficient intra-household allocations: A general characterization and empirical tests, *Econometrica*, *66*(6), 1241–1278.

Chiuri, M. C. (2000), Individual decisions and household demand for consumption and leisure, *Res. Econ.*, *54*, 277–324.

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

Dosman, D., and W. Adamowicz (2006), Combining stated and revealed preference data to construct an empirical examination of intrahousehold bargaining, *Rev. Econ. Household*, *4*, 15–34.

Dupont, D., W. L. Adamowicz, and A. Krupnick (2010), Differences in water consumption choices in Canada: The role of socio-demographics, experiences, and perceptions of health risks, *J. Water Health*, *8*(4), 671–686.

Ferrini, S., and R. Scarpa (2007), Designs with a-priori information for non-market valuation with choice-experiments: A Monte Carlo study, *J. Environ. Econ. Manage.*, *53*, 342–363.

Fiebig, D. G., M. P. Keane, J. Louviere, and N. Wasi (2010), The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity, *Marketing Sci.*, *29*(3), 393–421, doi:10.1287/mksc.1090.0508.

Genius, M., E. Hatzaki, E. M. Kouromichelaki, G. Kouvakis, S. Nikiforaki, and K. P. Tsagarakis (2008), Evaluating consumers' willingness to pay for improved potable water quality and quantity, *Water Resour. Manage.*, *22*, 1825–1834.

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

Hensher D. A., N. Shore, and K. Train (2005), Households' willingness to pay for water service attributes, *Environ. Resour. Econ.*, *32*, 509–531.

Hensher, D. A., S. M. Puckett, and J. M. Rose (2007), Extending stated choice analysis to recognise agent-specific attribute endogeneity in bilateral group negotiation and choice: A think piece, *Transportation*, *34*(6), 667–679.

Hensher, D. A., J. M. Rose, and I. Black (2008), Interactive agency choice in automobile purchase decisions: The role of negotiation in determining equilibrium choice outcomes, *J. Transport Econ. Policy*, *42*(2), 269–296.

Hensher, D. A., M. J. Beck, and J. M. Rose (2011), Accounting for preference and scale heterogeneity in establishing whether it matters who is interviewed to reveal household automobile purchase preferences, *Environ. Resour. Econ.*, *49*, 1–22.

Hess, S., K. E. Train, and J. W. Polak (2006), On the use of a modified Latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice, *Transport. Res. B*, *40*, 147–167.

Kanninen, B. (1993), Optimal experimental design for double-bounded dichotomous choice contingent valuation, *Land Econ.*, *69*(2), 138–146.

Keane, M. (2006), The generalized logit model: Preliminary ideas on a research program, presentation at Motorola-CenSoC Hong Kong Meeting, 22 October.

Kerr, G. N., and B. M. H. Sharp (2010), Choice experiment adaptive design benefits: A case study, *Aust. J. Agr. Resour. Econ.*, *54*, 40–420.

Lampietti, J. (1999), Do husbands and wives make the same choices? Evidence from Northern Ethiopia, *Econ. Lett.*, *62*, 253–260.

Louviere, J., and T. Eagle (2006), Confound it! That pesky little scale constant messes up our convenient assumptions, Proceedings of the 2006 Sawtooth Software Conference, pp. 211–228, Sawtooth Software, Sequem, Washington.

Lusk, J. L., and F. B. Norwood (2005), Effect of experimental design on choice-based conjoint valuation estimates, *Am. J. Agr. Econ.*, *87*(3), 771–785

Marcucci, E., A. Stathopoulos, L. Rotaris, and R. Danielis (2011), Comparing single and joint preferences: A choice experiment on residential location in three-member households, *Environ. Plan. Part A*, *43*(5), 1209–1225.

McFadden, D., and K. Train (2000), Mixed MNL models for discrete response, *J. Appl. Econ.*, *15*(5), 447–470.

Puckett, S. M., D. A. Hensher, J. M. Rose, and A. Collins (2007), Design and development of a stated choice experiment for interdependent agents: Accounting for interactions between buyers and sellers of urban freight services, *Transportation*, *34*(4), 429–451.

Quiggin, J. (1998), Individual and household willingness to pay for public goods, *Am. J. Agr. Econ.*, *80*, 58–63.

Rose, J. M., and M. C. J. Bleimer (2009), Stated preference experimental design strategies, *Transport Rev.*, *29*(5), 587–617.

Sandor, Z., and M. Wedel (2001), Designing conjoint choice experiments using managers' prior beliefs, *J. Marketing Res.*, *38*(4), 430–444,

Scarpa, R., and J. M. Rose (2008), Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why, *Aust. J. Agr. Resour. Econ.*, *52*, 253–282.

Scarpa, R., K. G. Willis, and M. Acutt (2007), Individual-specific welfare measures for public goods: A latent class approach to residential customers of Yorkshire Water, in *Econometrics Informing Natural Resource Management*, edited by P. Koundouri, pp. 316–337, Edward Elgar Publisher, Cheltenham, U. K.

Scarpa, R., M. Thiene, and K. Train (2008), Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps, *Am. J. Agr. Econ.*, *90*(4), 994–1010.

Smith, V., and G. V. Houtven (1998), Non-market valuation and the household, Discussion Paper 98-31, 1998 Resources for the Future.

Strand, J. (2007), Public-good valuation and intra-family allocation, *Environ. Resour. Econ.*, *38*, 527–543.

Swait, J., and J. Louviere (1993), The role of the scale parameter in the estimation and use of multinomiallogit models, *J. Marketing Res.*, *30*, 305–314.

Thiene, M., and R. Scarpa (2009), Deriving and testing efficient estimates of WTP distributions in destination choice models, *Environ. Resour. Econ.*, *44*(3), 379–395.

Train K. (1998), Recreation demand models with taste differences over people, *Land Econ.*, *74*(2), 230–239.

Train, K., and M. Weeks (2005), Discrete choice models in preference space and willing-to-pay space, in *Applications of Simulation Methods in Environmental and Resource Economics*, edited by R. Scarpa and A. Alberini, pp. 1–16, Springer, Dordrecht, The Netherlands.

Vermeulen, B., P. Goos, R. Scarpa, and M. Vandebroek (2011), Conjoint choice designs for measuring WTP, *Environ. Resour. Econ.*, *48*, 129–149.

Vermeulen, F. (2002), Collective household models: Principles and main results, *J. Econ. Surv.*, *16*, 533–564.

Willis, K., R. Scarpa, and M. Acutt (2005), Assessing water company customer preferences and willingness to pay for service improvements: A stated choice analysis, *Water Resour. Res. 41*, W02019.

D. A. Hensher, Institute of Transport and Logistics Studies, The Business School, University of Sydney, 144 Burren St., Newtown Campus, Sydney, NSW 2006, Australia. (david.hensher@sydney.edu.au)

R. Scarpa, Department of Economics, Waikato Management School, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. (rscarpa@waikato.ac.nz)

M. Thiene, Dipartimento Territorio e Sistemi Agroforestali Agripolis, Via dell'Università 16, I-35020 Legnaro (PD), Italy. (mara.thiene@unipd.it)