COLLEGIUM OF ECONOMIC ANALYSIS
WORKING PAPER SERIES

Estimating the membership function of the fuzzy willingness-to-pay/accept for health via Bayesian modelling

Michał Jakubczyk

Estimating the membership function of the fuzzy willingness-to-pay/accept for health via Bayesian modelling

Michał Jakubczyk
Decision Analysis and Support Unit
Warsaw School of Economics
Warsaw, Poland
michal.jakubczyk@sgh.waw.pl, http://michaljakubczyk.pl

***Abstract*—Determining how to trade off individual criteria is often not obvious, especially when attributes of very different nature are juxtaposed, e.g. health and money. The difficulty stems both from the lack of adequate market experience and strong ethical component when valuing some goods, resulting in inherently imprecise preferences. Fuzzy sets can be used to model willingness-to-pay/accept (WTP/WTA), so as to quantify this imprecision and support the decision making process. The preferences need then to be estimated based on available data. In the paper I show how to estimate the membership function of fuzzy WTP/WTA, when decision makers' preferences are collected via survey with Likert-based questions. I apply the proposed methodology to an exemplary data set on WTP/WTA for health. The mathematical model contains two elements: the parametric representation of the membership function and the mathematical model how it is translated into Likert options. The model parameters are estimated in a Bayesian approach using Markov-chain Monte Carlo. The results suggest a slight WTP-WTA disparity and WTA being more fuzzy as WTP. The model is fragile to single respondents with lexicographic preferences, i.e. not willing to accept any trade-offs between health and money.**

I. Introduction

Decision making with multiple criteria requires, explicitly or implicitly, making trade-offs between attributes describing decision alternatives. Even if the criteria are quantifiable and expressed as numbers (not only as labels along nominal or ordinal scale, e.g. *ugly*, *mediocre*, and *beautiful*), they may be of very different type, making it difficult to juxtapose them and decide about the exact trade-off coefficient. This is the case when non-market goods, such as health, safety, clean environment, etc., are being valued against money. The present paper focuses on juxtaposing health and financial consequences.

The amount of health gained with a given decision can, in principle, be expressed as a number: an increase in the life expectancy or—more generally—the additional quality-adjusted life years (QALYs). The latter combines the improvements in quality and longevity of life, is formally founded in axiomatic approach [1], and is operationally calculated via assigning numerical values, von Neumann-Morgenstern utilities, to health states [2] defined within some system, e.g. EQ-5D-3L [3]. Still, it is difficult to put a precise monetary value on health, as strikingly visible in systematic reviews of published estimates of the value of statistical life: the standard deviations of published results (e.g. within a given country) are usually of the order of magnitude of the mean values [4], [5], [6]. The valuation of non-market goods is also specific in a sense that there is a great difference between the willingness-to-pay (WTP, the amount one is willing to pay for an additional unit of good) and the willingness-to-accept (WTA, the amount one demands to obtain in order to accept the loss of one unit) [7]. In spite of these difficulties, it is necessary to grasp the preferences quantitatively in order to support the decision making process and make it transparent.

In social sciences, in order to formally model imprecision, fuzzy sets and fuzzy logic have been used for many years [12]. Therefore, as a solution, it was suggested to also treat the WTP and WTA for health as a fuzzy concept [8], i.e. to define a fuzzy number fWTP (or fWTA) over the universe of $\mathbf{R}_+$ with a non-increasing (non-decreasing, respectively) membership function $\mu_{\text{fWTP}}(\lambda)$ ($\mu_{\text{fWTA}}(\lambda)$), interpreted as the conviction that it makes sense to pay (accept) λ for an additional QALY (a loss of QALY). It was shown how to support decision making via fuzzy preference relations [8] or choice functions, when multiple alternatives are present [9], [10]. One of the choice functions used the 0.5-cut of fWTP/fWTA, and formal statistical methods how to infer this value based on data collected via surveys in random samples were shown [10]. Some choice functions, however, required knowing the complete shape of the membership function, and learning this shape allows, additionally, to better understand the nature of imprecision in the perception of WTP/WTA, e.g. to compare the amount of fuzziness between these two. Building on the above motivation, in the present paper I show how to estimate the membership functions using the data collected via Likert-based surveys in random samples.

II. Data

I use the data set previously described in the literature [8]. Briefly, 27 health technology assessment experts in Poland were asked to express their views on how much a society should be willing to pay (should demand to be compensated) for an additional QALY (for a QALY lost). Importantly, the experts were asked to think about the societal value, i.e. the value that should be used in public decision making, not about how much they value the improvement/worsening in their own lives. Therefore, all the respondents were asked about the same thing; while still allowing for a difference in opinions. The respondents were asked about their personal views, not simply to quote the current regulations, which in Poland define the threshold to be precisely three times the annual gross domestic product per capita.

In the survey the respondents were asked, inter alia, if they would accept using a technology offering one QALY more (less) if it costed λ more (less). The respondents declared how much they agreed, using a 5-level Likert scale: 5—totally agree, 1–totally disagree. The raw collected responses are illustrated in Figure 2. The exact λ values can be read off the horizontal axis ($1≈4 Polish Zloty, PLN).

Three respondents were altogether removed in the WTA part (but included in WTP analysis and in sensitivity analysis), as they did not use the *tend to agree* & *agree* Likert answers for any, even strikingly large λ. Thus, they seemed to disagree with the very possibility of trading off health for money, in a sense rejecting the rules of the game. This paper focuses on estimating the membership function of WTP/WTA, and these three experts reject the very concept of being willing to accept money for health, hence their views cannot be used to quantitatively estimate (crisp or fuzzy) WTA. Still, these opinions are important in constructing a general framework how to decide about public spending in healthcare (perhaps rejecting the very idea of explicit trade-offs), but should be handled separately as they differ qualitatively.

To motivate the present research, let us notice here that—on one hand—it would be, in principle, possible to estimate the membership function in a naïve approach, as follows. We could translate the Likert levels 1–5 into values $\{0, 0.25, 0.5, 0.75, 1\}$, respectively, and average the resulting values (separately for each λ) between the respondents. Then we would simply interpret the resulting average as the value of the membership function, additionally somehow interpolating between available λs. On the other hand, such an approach has several disadvantages. Most importantly, averaging the membership function can result in the outcome being fuzzy, even without any fuzziness in the individual data. This is illustrated in Figure 1. Assume that four respondents have different, yet crisp, opinions about what values, x, belong to a given set, X, over a universe $\mathbf{R}$. In their view, respectively, $X = [0, 1]$, $X = [0, 2]$, $X = [0, 3]$, and $X = [0, 4]$. The membership functions are thus discontinuous (left panel) and take only values $\{0, 1\}$ (and the respondents would only use Likert levels 5 and 1, in their answers). Averaging the results yields a stepwise function (right panel), taking on also values between 0 and 1, and hence implying fuzziness. Thus a stochastic noise is transformed into fuzziness, while the two types of uncertainty are usually treated as qualitatively different.

Secondly, the naïve approach does not allow for any extrapolation, and so we would only estimate the membership function up to the maximal λ used in the questionnaire. Thirdly, this approach does not allow (easily) to model the heterogeneity of the respondents, so as to estimate the impact of some personal traits on the WTP/WTA. In principle the modelling approach presented below can include additional explanatory variables characterizing individuals to see how they are associated with the preferences.

Additionally, the naïve averaging might not work if the respondents were presented different λs in the questionnaire; and using various λs might be a good idea if we wanted to collect data for many distinct λs (e.g. to verify the impact of using round numbers), while not asking a single respondent too many questions. In the naïve approach, the averaging would be done over different subsets of respondents and might lead, e.g., to an increasing estimated $\mu_{\text{WTP}}(\cdot)$, a logical impossibility.

Finally, explicitly modelling how the individual opinions are translated into Likert levels (as equation 2 below) can allow combining various data in a flexible way and may help designing questionnaires in the future (e.g. how many levels to use).

III. Model

A. Mathematical formulation

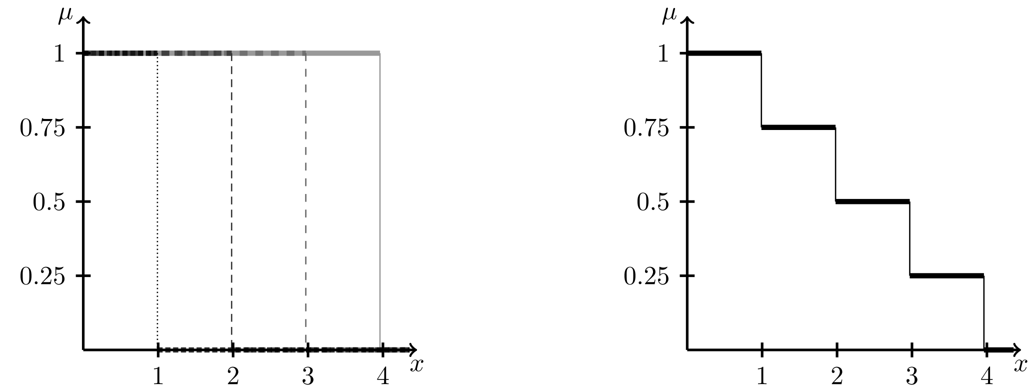
Below I present the model for WTP but the idea is analogous for WTA. In what follows, I assume that every individual has their own membership function, $\mu_i(\cdot)$ (WTP suppressed in the subscript, for brevity). I assume that this function is given parametrically by the following equation

$$\mu_i(\lambda) = \frac{1}{1 + a_i \lambda^{b_i}}, \tag{1}$$

where a_i and b_i are strictly positive parameters, specific to a given individual. Hence, I use a standard, decreasing S-shape logistic curve for the natural logarithm of λ. S-shape is to represent the smooth transition between the full and lack of conviction for greater and greater values (and the location and steepness is given as a function of the parameters, a and b, and so can differ between the respondents). Using the logarithm of λ reflects diminishing sensitivity to equal, absolute increases in cost difference, is motivated by the skewness found in the data [10], and makes the approach robust to considering PLNs per QALY or QALYs per PLN. Still, non-log values are used in sensitivity analysis.

I then assume that, when the respondent faces a survey, $\mu_i(\lambda)$ is translated into the Likert levels $L_i(\lambda)$ probabilistically, i.e. for each $\mu_i(\lambda)$ there is a probability distribution defined over levels 1–5. Intuitively, the probability of observing a specific level $L_i(\lambda) = k$, $k = 1, \ldots, 5$, being selected diminishes the farther away $\mu_i(\lambda)$ is from this level's threshold value, θ_k, $\theta_5 = 1$, $\theta_4 = 0.75$, $\theta_3 = 0.5$, $\theta_2 = 0.25$, and $\theta_1 = 0$. More formally,

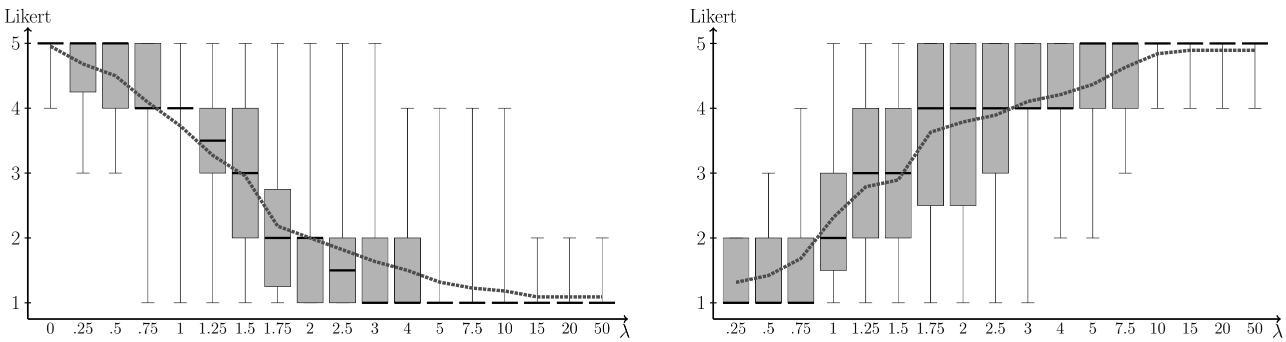
$$P\left(L_i(\lambda) = k\right) = \frac{\exp(-s_i \times |\mu_i(\lambda) - \theta_k|)}{\sum_{j=1}^{5} \exp(-s_i \times |\mu_i(\lambda) - \theta_j|)}. \tag{2}$$



(a) Individual data (various types of lines used for individual respondents)

(b) Result of averaging

Fig. 1. The artefact of naïve averaging of the membership function, μ. Individuals have crisp, however different, opinions. The average membership function takes on values between 0 and 1, suggesting fuzziness. Discontinuities of μ denoted with thin, vertical lines.



(a) WTP

(b) WTA

Fig. 2. Data. Values in horizontal axis in hundreds 000s PLN ($\lambda = 0$ was not used in WTA in the survey). Bold line denotes the median. Gray box denotes the first and third quartile. The whiskers denote the min and max. The dotted lines denote the naïvely calculated average response.

Parameter s_i, $s_i > 0$, measures the diffusion of probabilities along the Likert-scale. Figure 3 illustrates the above equation in work for three values of s. Small s_i results in the probabilities being distributed more evenly, and so for any μ_i still all the Likert levels are quite probable (e.g. the respondent is not capable of perceiving own μ_i or is not answering the Likert questions meticulously enough). Increasing s_i results in the level with θ closest to $\mu_i(\lambda)$ being selected with increasing probability. Thus, importantly, the present specification includes as a special case (with $s_i \to +\infty$) selecting the Likert answer closest to $\mu_i(\lambda)$ (when interpreting Likert answers being transformed to 0, 0.25, 0.5, 0.75, and 1).

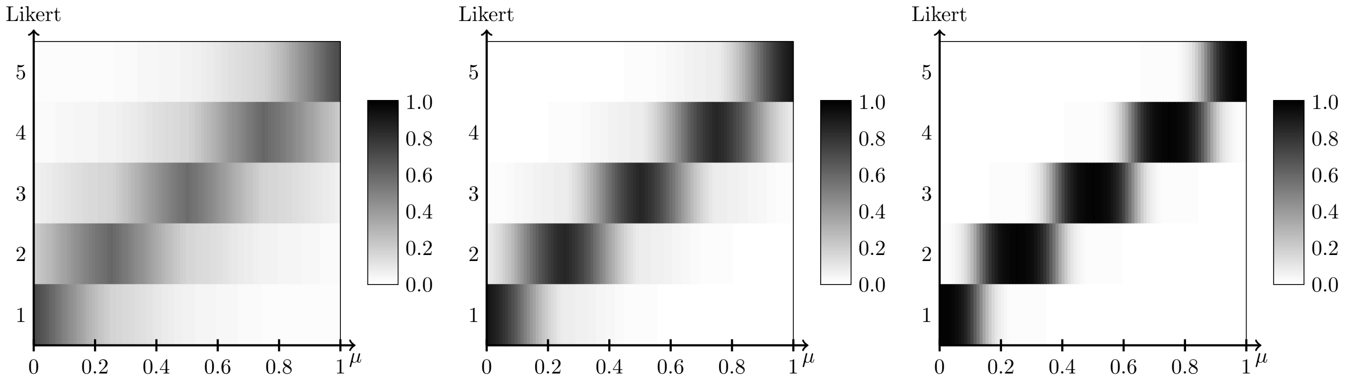
Another nice (in my view) feature of the above approach is that, e.g., even if $\mu_i(\lambda) = 0.76$, it is still possible for the middle option (and any other) to be selected, as the numerator in equation 2 is always strictly positive. In that sense, the respondent is not entirely able to perceive own membership function being equal to 0.76 so as to rule out the middle option due to it being blocked by level 4, with threshold value 0.75. Of course, other formulas could be used, e.g. implying randomizing only between two levels, directly above and below a given $\mu_i(\lambda)$.

Parameters a_i, b_i, and s_i are idiosyncratic for each respondent. They are drawn independently—of each other, and between respondents—from lognormal distributions with the underlying normal distributions with means and inverse variances $N(m_A, \tau_A)$, $N(m_B, \tau_B)$, and $N(m_S, \tau_S)$, respectively. Point estimates $\widehat{m_A}$, $\widehat{m_B}$, and $\widehat{m_S}$ are taken to define the estimand, i.e. population-level membership function.

The point estimates are taken as medians of posterior distributions. Percentiles 2.5 and 97.5 define the 95% credible interval (95%CI). On technical note, non-informative priors are assumed (normal for means, gamma for inverse variances). The model is estimated with Markov-Chain Monte Carlo method, implemented in JAGS/R. The results come from 20,000 iterations (thinning=5), with 5000 burn-in iterations.

B. JAGS code

Below I present the JAGS code used to estimate the model. The following notation is used. `nR` and `nV` denote the number of respondents and λs, respectively. `answer[i,j]` is a matrix of Likert answers, `i` and `k` indexing respondents and λs, respectively, `l[k]` $= [0, 0.25, 0.5, 0.75, 1]$. `mA`, `mB`, and `mS` are the estimated m_A, m_B, and m_S, respectively.



(a) 5 (b) 10 (c) 20

Fig. 3. The conversion of a membership function (horizontal axis) into probabilities of Likert levels (shades of gray for various levels on vertical axis) for three values of diffusion parameter, s_i.

```
model{
#OUTER LOOP OVER RESPONDENTS:
for (i in 1:nR){

 #DRAW INDIVIDUALS' PARAMETERS:
 a[i] ~ dlnorm(mA,tA)
 b[i] ~ dlnorm(mB,tB)
 s[i] ~ dlnorm(mS,tS)

 #MIDDLE LOOP OVER LAMBDAS:
 for (j in 1:nV){
  x[i,j] <- 1/(1+a[i]*pow(v[j],b[i])) #***

  #INNER LOOP OVER LIKERT LEVELS:
  for (k in 1:5){
   #PROB. OF LEVEL K (NOT NORMALIZED)
   p[i,j,k] <- exp(-s[i]*abs(x[i,j]-l[k]))
  }

  #PROB. DIST. OF OBSERVABLES:
  answer[i,j] ~ dcat(p[i,j,1:5])
 }
}

#PRIORS
mA ~ dnorm(0,.01)
mB ~ dnorm(0,.01)
mS ~ dnorm(0,.01)
tA ~ dgamma(.01,.01)
tB ~ dgamma(.01,.01)
tS ~ dgamma(.01,.01)
}
```

In order to use the code for WTA, it suffices to replace the $\star\star\star$-marked line with

```
x[i,j] <- 1-1/(1+a[i]*pow(v[j],b[i]))
```

as we expect an increasing membership function, and the sign of parameters is predetermined by lognormal distributions.

IV. Results

A. Baseline results

The estimation yields the following results for WTP:

- $\widehat{\mu_A} = -17.62$ with $95\%\text{CI} = (-20.95; -14.13)$,
- $\widehat{\mu_B} = 1.25$ with $95\%\text{CI} = (0.99; 1.44)$,
- $\widehat{\mu_S} = 2.32$ with $95\%\text{CI} = (2.06; 2.61)$.

For the sake of estimating the 0.5-cut, the above implies that $\mu = 0.5$ for ca. 155,700 PLN/QALY, while the official threshold amounts currently to 125,955 PLN/QALY (and 111,381 PLN/QALY at the time of the survey).

The results for WTA are as follows:

- $\widehat{\mu_A} = -15.43$ with $95\%\text{CI} = (-19.58; -11.65)$,
- $\widehat{\mu_B} = 1.03$ with $95\%\text{CI} = (0.74; 1.33)$,
- $\widehat{\mu_S} = 2.48$ with $95\%\text{CI} = (2.28; 2.7)$,

and we get $\mu = 0.5$ for ca. 246,800 PLN/QALY. Hence, there is some disparity in 0.5-cuts. We still have to take into consideration that the current methodology was not focused on estimating the 0.5-cut, but on the whole membership function, and so the estimates of the 0.5-cut may be driven by other regions of the membership function.

The resulting membership functions, based on point estimates, are presented in Figure 4 (the left panel). Now, having estimated the membership functions allows comparing WTP and WTA quantitatively. Firstly, notice that the 95%CI are wider for WTA (for μ_A and μ_B, only which matter for the shape of the membership function), and so there is more stochastic uncertainty related to WTA than WTP. We can also compare the two with respect to the amount of fuzziness. I use two measures for that purpose, one was proposed in the literatures and relies on the idea that the more the membership function takes on values close to 0.5, the fuzzier the set is [11]. Formally, the amount of fuzziness is given by

$$\int_{-\infty}^{+\infty} (1 - |2\mu(x) - 1|)\ \mathrm{d}x.$$

We obtain, for WTP, 137.98, and for WTA, 284.87 (calculated numerically). I use another approach here, using the particular shape of the membership function (decreasing monotonically from 1 to 0). We can interpret the membership function in the probabilistic fashion, as a flipped vertically cumulative distribution function, and calculate the variance of the variable, λ. Then the more the (upper bounds of the) α-cuts differ, the fuzzier the set is. More formally, the amount of fuzziness is then given by

$$\sqrt{\int_0^1 \sup(\mathrm{WTP}_\alpha)^2 \mathrm{d}\alpha - \left(\int_0^1 \sup(\mathrm{WTP}_\alpha)\mathrm{d}\alpha\right)^2}.$$

We obtain 108.07 for WTP and 253.44 for WTA, and so again—there is twice as much imprecision (in a fuzzy set theory sense) in how WTA is perceived. Probably experts feel more uncomfortably with quantifying trade-offs in the context of selling health, which results in more of both—stochastic noise (differences between the respondents) and fuzziness (inability to precisely locate the threshold).

B. Results for WTA for a complete data set

As a sensitivity analysis, I repeat the calculations for the complete data set, i.e. keeping in the three respondents that were removed for WTA analysis. Then, only the results for WTA change, as follows:

- $\widehat{\mu_A} = -15.13$ with 95%CI $= (-19.94; -10.7)$,
- $\widehat{\mu_B} = 0.79$ with 95%CI $= (0.33; 1.23)$,
- $\widehat{\mu_S} = 2.59$ with 95%CI $= (2.39; 2.85)$.

The membership function is also illustrated in Figure 4 (the left panel, dashed line). As can be seen, including the—qualitatively different-respondents results in the results being completely different, and not too reliable (taking into account the raw data, Figure 2). This stresses the necessity to handle the data violating the underlying assumptions (accepting some trade-offs) separately.

C. No logarithmic transformation

For baseline analysis log of λs were used. Here I verify the impact of this transformation, redoing the calculations for original values. For that purpose, instead of equation 1, we need to use the following one:

$$\mu_i(\lambda) = \frac{1}{1 + e^{a_i + b_i \times \lambda}}. \quad (3)$$

The above expression obviously simplifies to equation 1 when we take $\ln(\lambda)$ instead of λ (and denote e^{a_i}, abusing notation, by a_i). We also need to make adequate changes in JAGS code in the line marked with $\star\star\star$, taking:

```
x[i,j] <- 1/(1+exp(a[i]+v[j]*b[i]))
```

Now, a_i can also take negative values, and so I assume it is normally distributed, with parameters $N(m_A, \tau_A)$.

The estimation yields, for WTP:

- $\widehat{\mu_A} = -4.2$ with 95%CI $= (-5.3; -3.24)$,
- $\widehat{\mu_B} = -3.68$ with 95%CI $= (-4.09; -3.28)$,
- $\widehat{\mu_S} = 2.5$ with 95%CI $= (2.29; 2.74)$,

and for WTA:

- $\widehat{\mu_A} = -3.92$ with 95%CI $= (-5.14; -2.88)$,
- $\widehat{\mu_B} = -3.92$ (coincidentally equal to $\widehat{\mu_A}$) with 95%CI $= (-4.42; -3.41)$,
- $\widehat{\mu_S} = 2.49$ with 95%CI $= (2.21; 2.87)$.

The results are depicted in the right panel of Figure 4. As can be seen, this approach reduces the WTP-WTA disparity in terms of 0.5-cuts. The problem with the non-log approach with the $\mu(\cdot)$ given by equation 3 is that now necessarily $\mu(0) < 1$ for WTP and $\mu(0) > 0$ for WTA, which does not seem intuitive and desirable. That's why this is not treated as the baseline result.

D. Results for WTA for non-log, complete data

For completeness, I present the results of the non-log approach with all the respondents in the WTA analysis. The estimation yields:

- $\widehat{\mu_A} = -3.84$ with 95%CI $= (-5.19; -2.75)$,
- $\widehat{\mu_B} = -4.56$ with 95%CI $= (-5.48; -3.68)$,
- $\widehat{\mu_S} = 2.7$ with 95%CI $= (2.33; 3.25)$.

The results are illustrated with the dashed line in the right panel of Figure 4. Apparently, the non-log approach is more robust to including respondents strongly opposing to accepting trade-offs resulting in losing some health.

V. Conclusion

Understanding the shape of the membership function of fuzzy trade-off coefficients is important to formally support decision making. When imprecise opinions about possible values of this coefficient are collected via surveys from a group of respondents, these judgements will most certainly differ, and statistical methods must be employed to estimate the joint, average one. Naïve methods may be misleading: e.g. they can artificially suggest more fuzziness than actually present in the raw data. Hence, approaches based on modelling—as one presented in the present paper—may be more useful.

The proposed model clearly distinguishes between the parameterisation of the membership function and the mechanism of its conversion into Likert scale, thus providing flexibility on changing the two independently. Respondents' characteristics could be added to equation 1 (or 2, or 3), thus allowing to model heterogeneity and find factors associated with, e.g., accepting greater WTP or larger WTP-WTA disparity.

Further work should be focused on building more robust model, that could handle also respondents with qualitatively different opinions (not crossing the middle Likert option). That may require using models with more flexibility to vary the membership function between the respondents, and so with more parameters: hence, larger data sets are required. As an idea: a scaling parameter could be used to limit the range of values that the membership function can take for a given respondent. Also, perhaps differently shaped functions should be tested for WTP and WTA. That could account for larger fuzziness for WTA, while restoring the equality of 0.5-cuts, detected by different methods [10].

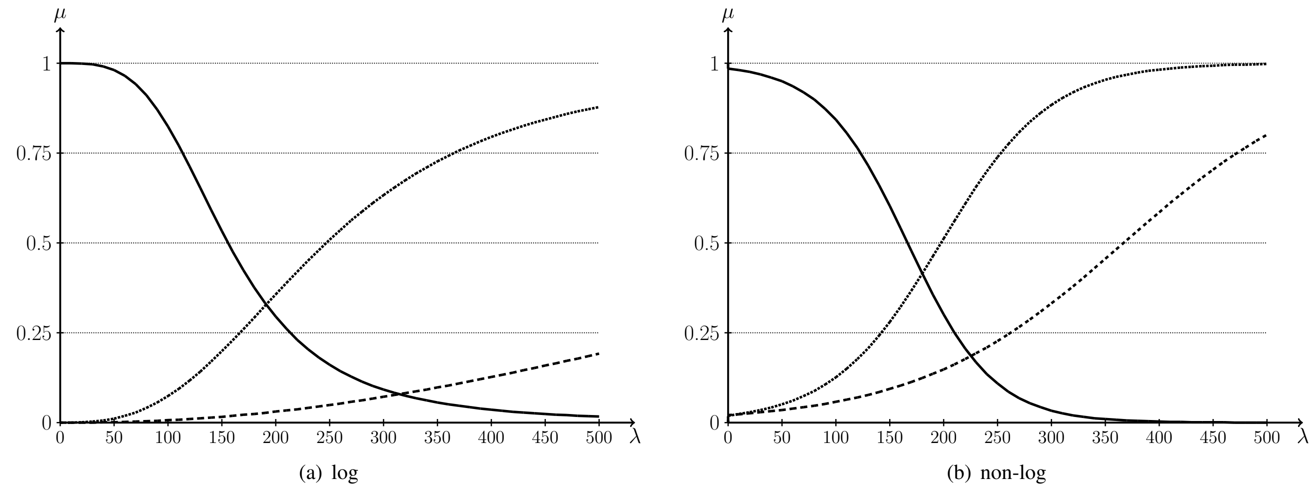


Fig. 4. Membership functions for point estimates (WTP—solid, WTA—dotted). Additionally, results for WTA (dashed) when no respondents are removed.

ACKNOWLEDGMENT

The research was done during my stay at The Tippie College of Business, The University of Iowa, USA, thanks to the Fulbright Senior Award. This opportunity is greatly appreciated.

REFERENCES

[1] H. Bleichrodt, P. Wakker, and M. Johannesson, "Characterizing QALYs by Risk Neutrality," *Journal of Risk and Uncertainty*, vol. 15, pp. 107–114, 1997.

[2] D. Golicki, M. Jakubczyk, M. Niewada, W. Wrona, and J. Busschbach, "Valuation of EQ-5D Health States in Poland: First TTO-Based Social Value Set in Central and Eastern Europe," *Value in Health*, vol. 13, pp. 289–297, 2010.

[3] R. Brooks and F. De Charro, "EuroQol: The current state of play," *Health Policy*, vol. 37, pp. 53–72, 1996.

[4] F. Bellavance, G. Dionne, and M. Lebeau, "The value of a statistical life: a meta-analysis with a mixed effects regression model," *Journal of Health Economics*, vol. 28, no. 2, pp. 444–464, 2009.

[5] W. Viscusi and J. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," *Journal of Risk and Uncertainty*, vol. 27, no. 1, pp. 5–76, 2003.

[6] L. Hultkrantz and M. Svensson, "The value of a statistical life in Sweden: a review of the empirical literature," *Health Policy*, vol. 108, no. 2–3, pp. 302–310, 2012.

[7] W. Dubourg, M. Jones-Lee, and G. Loomes, "Imprecise Preferences and the WTP-WTA Disparity," *Journal of Risk and Uncertainty*, vol. 9, pp. 115–133, 1994.

[8] M. Jakubczyk and B. Kamiński, "Fuzzy approach to decision analysis with multiple criteria and uncertainty in health technology assessment," *Annals of Operations Research*, doi: 10.1007/s10479-015-1910-9, 2015.

[9] M. Jakubczyk, "Using a fuzzy approach in multi-criteria decision making with multiple alternatives in health care," *Multiple Criteria Decision Making*, forthcoming, 2016.

[10] ——, "Choosing from multiple alternatives in cost-effectiveness analysis with fuzzy willingness-to-pay/accept and uncertainty," mimeo, 2016.

[11] G. Klir and T. Folger, *Fuzzy sets, uncertainty, and information*. Prentice-Hall, 1988.

[12] M. Smithson, *Fuzzy Set Analysis for Behavioral and Social Sciences*. Springer-Verlag, 1987.