Use of a nonparametric Bayesian method to model health state preferences: An application to UAE EQ-5D-5L valuations

Samer A Kharroubi [1, 2], PhD, Cstat, Fatima Al Sayah[3], PhD

[1] Department of Nutrition and Food Sciences, American University of Beirut, Beirut, Lebanon
[2] Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield, Sheffield, UK
[3] School of Public Health, University of Alberta, Edmonton, AB, Canada

1. Introduction

This paper reports on the findings from applying a Bayesian approach to modelling health-related quality of life (HRQoL) data. The approach applies a nonparametric model to estimate EQ-5D-5L health state utility values using Bayesian methods. The data set is the UAE EQ-5D-5L valuation study where a sample of 86 health states defined by the SF-6D was valued by a sample of 1005 UAE general population using standard gamble. The paper presents the results from applying the nonparametric model and comparing it to the original model estimated using a conventional parametric random effects. The two models are compared theoretically and in terms of empirical performance. The nonparametric Bayesian model is a powerful technique for analyzing HRQoL data and is argued to be more theoretically appropriate than previously used parametric models.

2. Highlights

- The paper demonstrates that the nonparametric Bayesian method is a superior modelling choice, with the potential ultimate goal being to encourage researchers to use this modelling choice over a parametric approach.
- The superiority of the non-parametric Bayesian approach may prove to be important in practice and hence highly likely to be the first choice for applied researchers.

3. Methods

Descriptive System

The EQ-5D-5L includes a descriptive system and the EuroQol visual analog scale (EQ-VAS) [1]. The descriptive system assesses 5 dimensions of health: *mobility, self-care, usual activities, pain/discomfort, and anxiety/depression*. Each dimension has 5 levels of severity—*no, mild, moderate, severe, and extreme/unable to*—describing 3125 distinct health states. Represented by a 5-digit number, these health states range from 11111 (perfect health) to 55555 (worst health state). The EQ-VAS captures the patient's self-rated health on a scale ranging from 0 "the worst health you can imagine" to 100 "the best health you can imagine."

Study design

This study followed the EuroQol valuation protocol for the EQ-5D-5L [2,3]. Using the EuroQol Valuation Technology (EQ-VT) platform, we conducted a computer-based, interviewer-administered face-to-face or online survey with a representative sample of the general UAE population.

Preference Elicitation Methods and Health State Selection

The EuroQol EQ-5D-5L valuation protocol uses 2 preference elicitation methods: composite time trade-off (cTTO) and discrete choice experiment (DCE) [2,3]. Briefly, the cTTO method evaluates an EQ-5D-5L health state in relation to full health. It uses conventional time trade-off (TTO) to determine values for health states perceived as better than dead and leadtime TTO for states viewed as worse than dead. Afterward, respondents complete the DCE tasks where they express preferences between 2 different EQ-5D-5L health state profiles in pairwise comparisons

Sampling and Recruitment Strategy

A total of 1150 adult UAE participants were targeted. Recruitment used a 2-stage quota sampling strategy. First, we stratified geographically by Emirate, and then strata were defined based on age and sex distribution of the general population [4].

Data Collection

Data collection was conducted between January and August 2023. The interview process consisted of the following steps: (1) interviewer explained the purpose of the study, its procedure, confidentiality, and risk/benefits to the participant; (2) participant completed the EQ-5D-5L; (3) interviewer explained (using the wheelchair example) and administered the cTTO task (10 health states); (4) interviewer explained and administered the DCE task (7 pairs); and (5) participant completed a country-specific questionnaire including questions on sociodemographics, health history, and feedback on TTO and DCE tasks.

4. Modelling

The parametric approach

A general parametric model for health state valuations can be specified as:

$$y_{ij} = \mu + \theta' g(\mathbf{x}_{ij}, \boldsymbol{t}_j) + \alpha_j + \varepsilon_{ij} \qquad (1)$$

where, for $i = 1, \dots, n_j$ and $j = 1, \dots, m$, $\mathbf{x}_{ij}$ is the i^{th} health state valued by individual j and y_{ij} is the adjusted SG value obtained by individual j for that health state. Further, μ and θ represent unknown parameters, $g(\mathbf{x}_{ij}, \boldsymbol{t}_j)$ is a vector of dummy explanatory variables and $\boldsymbol{t}_j$ denotes individual characteristics. The model also includes a zero-mean random terms: the individual effect term α_j of respondent j and the error term ε_{ij}.

The nonparametric approach

The i^{th} valuation by individual j is modelled as [5]:

$$y_{ij} = 1 - \alpha_j\{1 - u(\mathbf{x}_{ij})\} + \varepsilon_{ij} \qquad (2)$$

where

$$\alpha_j \sim LN(\boldsymbol{t}_j^T \gamma, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, v^2)$$

$\mathbf{t}_j$ denotes a vector of covariates for individual j, and γ, τ^2 and v^2 are unknown parameters. The term $u(\mathbf{x})$ is an unknown function and represents the median utility for health state x. Therefore, in Bayesian inference, it becomes a random variable. The prior distribution of $u(\mathbf{x})$ is

$$u(\mathbf{x}) \sim N(\mathbf{h}(\mathbf{x})^T \beta, \sigma^2) \qquad (3)$$

where $\mathbf{h}(\mathbf{x}) = (1, \mathbf{x})^{\boldsymbol{T}}$, $\mathbf{x}$ is a vector of discrete levels on each of the six health dimensions and β and σ^2 are unknown parameters. Note that the mean function of (3) represents a belief that the utility will be approximately linear and additive in the different dimensions, but whereas (1) imposes this linearity and additivity as a strict assumption about the utility function, our model expresses it as a prior expectation.

Further, we expect there to be a high correlation between $u(\mathbf{x})$ and $u(\mathbf{x}')$ if $\mathbf{x}$ and $\mathbf{x}'$ get sufficiently close. This is defined as

$$\text{cov}(u(\mathbf{x}), u(\mathbf{x}')|\sigma^2) = \sigma^2 c(\mathbf{x}, \mathbf{x}')$$

where

$$c(\mathbf{x}, \mathbf{x}') = \exp\{-\sum b_d(\mathbf{x}_d - \mathbf{x}'_d)^2\}$$

where for $d = 1, \dots, 5$, x_d and x'_d represent the levels of dimension d in $\mathbf{x}$ and $\mathbf{x}'$ respectively, and b_d is the roughness parameter that controls the degree of adherence of $u(\mathbf{x})$ to a linear form in dimension d [5].

Note that the mean health state utility in (2) is

$$\bar{u}(\mathbf{x}) = 1 - E(\alpha)\{1 - u(\mathbf{x})\}$$

where $E(\alpha)$ is mean value of α over the whole population. Note that, if $E(\alpha) = 1$, then $\bar{u}(\mathbf{x}) = u(\mathbf{x})$.

MATLAB code for implementing the Bayesian nonparametric model can be found at https://www.sheffield.ac.uk/scharr/sections/heds/mvh/sf-6d/bayesian.

5. Results

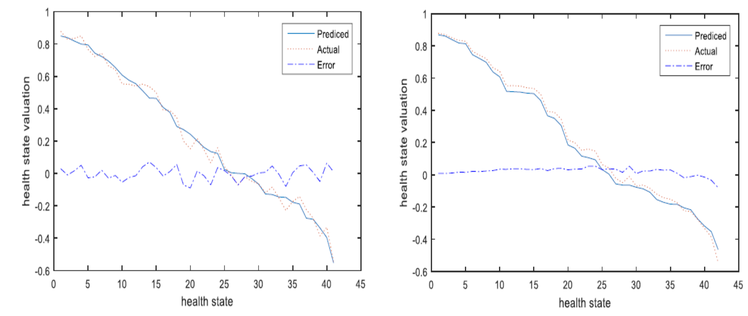
The models are compared in terms of their predictive ability in Figs. 1a and 1b, where the predicted and actual mean values for the 86 health states valued in the survey have been plotted with health states ordered by predicted values.

The plots suggest that the parametric model over predicts the value of the better states, whereas this does not seem to have been a problem for the nonparametric model. This provides an initial indication that the nonparametric model is less prone to systematic bias.

Across 86 health states, the predictive performance of model (2) is better than that of model (1) overall, with a root mean square error (RMSE) of 0.035 for model (2) compared to 0.062 for model (1).

Figure 1. Actual and predicted mean health states valuations for the (a) parametric model and (b) nonparametric model

(a) (b)



Out-of-sample leave-one-out prediction

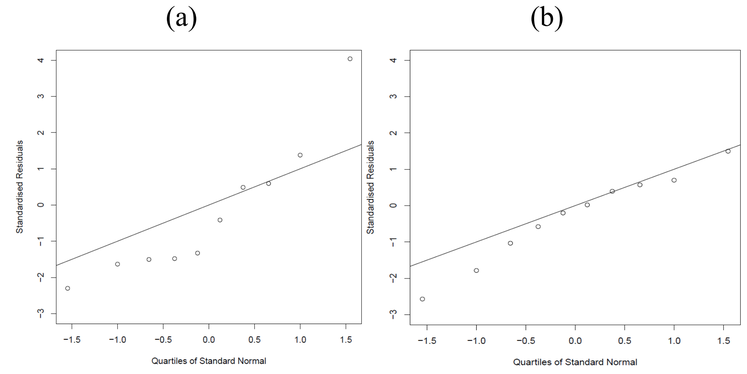
Data relating to 10 selected health states were removed randomly from the estimation data, and the models fitted on data for the remaining 76 states. The predictive performance of model (2) is better than that of model (1) overall, with RMSEs of 0.054 and 0.099.

Figs. 2a and 2b show the Q–Q plots of standardized predictive errors for the 10 health states sample means, for models (1) and (2), respectively. In each figure the straight line corresponds to the theoretical $N(0,1)$.

In Fig. 2a, the points deviate substantially from the theoretical line, therefore, the parametric model is not well validated by its predictive performance. In contrast, it is apparent from Fig. 2b that the nonparametric model predictions are well validated.

Figure 2. Q-Q plot of standardized predictive errors for the 10 out of sample health states for the (a) parametric model and (b) nonparametric model.

(a) (b)



6. Discussion

The nonparametric model has three principal advantages over the conventional parametric model.

- *Value of perfect health*: every respondent values perfect health as having utility 1, no matter what that respondent's covariates are and regardless of the random effect.
- *Respondent effects*: The covariate and respondent random effects enter multiplicatively rather than additively, thereby allowing for more effect the further the base utility is away from 1.
- *Flexibility*: The nonparametric model allows the preference function over health states to take any shape at all.

7. References and data sharing

References

1. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10): 1727–1736..
2. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, de Charro F. A programme of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17(4):445–453.
3. Oppe M, Norman R, Yang Z, van Hout B. Experimental design for the valuation of the EQ-5D-5L. In: Devlin N, Roudijk B, Ludwig K, eds. Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide. Cham: Springer; 2022.
4. Government U. UAE total population estimates by age group - gender. Official Data Portal of the UAE Government; Published 2023. Bayanat.ae.
5. Kharroubi SA, O'Hagan A, Brazier JE. Estimating Utilities from individual health state preference data: a nonparametric Bayesian approach. Applied Statistics 2005; 54: 879-895.

Ethics Statement

- The study was ethically approved by the Social Sciences Research Ethics Committee at the United Arab Emirates University (ERSC_2022_1648).