Bayesian Inference for health state utilities using pairwise comparison data

Theresa Cain

Submitted for the degree of PhD

School of Mathematics and Statistics

March 2011

University of Sheffield

PAGINATED
BLANK PAGES
ARE SCANNED AS
FOUND IN
ORIGINAL
THESIS

NO
INFORMATION
MISSING

Abstract

The National Institute for Health and Clinical Excellence (NICE) is responsible for making recommendations about which treatments are available on the NHS. An important part of the decision making process is to estimate the cost effectiveness of a treatment, measured in cost per QALY gained. If a treatment costs more than £30000 per QALY the NHS does not consider it to be cost effective. QALYs are calculated using life years and QALY weights, which represent the quality of life of a condition. An example of a QALY weight is a utility, which is a measure of preference for a health condition. A utility is measured on a scale between 0 and 1, where 0 is the utility of death and 1 is the utility of perfect health. This thesis uses discrete choice modelling to estimate utilities for health states defined using the Asthma quality of life questionnaire. A Bayesian approach is used to estimate the utilities in order to quantify utility. A probit and logit model are considered for the likelihood where the parameters represent the decrease in utility associated with increasing levels of the attributes of the asthma quality of life questionnaire. An MCMC is run using three prior distributions on the parameters: Gamma(1,10), Gamma(5,15) and Uniform(0,1). The model is also extended to include a multiplicative random effect. Bayes factors are used as a model comparison in the standard model. Results from both the standard model and random effects model are also compared with maximum likelihood estimates.

Contents

Chapter 1

Introduction

In health care there are many conditions that need to be treated and for each condition there may be several possible treatments. Health care providers, such as the NHS, have limited financial resources and therefore are not able to provide the best available treatment for every possible condition. Financial resources need to be allocated in an effective way in order to provide the best treatment for as many patients as possible.

The National Institute for Health and Clinical Excellence (NICE) is responsible for making recommendations to the NHS in England and Wales. These recommendations cover three areas: new and existing medicines, treatments and procedures; treating and caring for people with specific diseases and conditions; and how to improve health and prevent disease and illness. In order to determine recommendations on new and existing treatments, NICE undertakes a Health Technology appraisal at the request of the Department of Health. The process, described in NICE (2008), assesses the clinical and cost effectiveness of a health technology(treatment or drug) and is divided into three phases: scoping, assessment and appraisal.The scope defines the issues of interest and the questions that should be addressed. Assessment usually consists of two components: a system-

atic review and an economic evaluation. The am of the assessment process is to produce an assessment, taking into account uncertainty, of the clinical and cost effectiveness of a specific health technology. The appraisal process then considers the reports and analysis produced in the assessment phase within the context of additional information supplied by appropriate experts and the general public.

An important part of the assessment process of a Health Technology appraisal is the economic evaluation. Drummond et al. (2005) describe economic evaluation as the comparative analysis of alternative courses of action in terms of both their costs and consequences. A type of economic evaluation is cost-effectiveness analysis (CEA), which compares the cost of a treatment with its effectiveness, measured in natural units relevant to the condition being investigated (e.g. millimetres of blood pressure reduction). The results are expressed as cost per unit of effectiveness.

There are several problems associated with the use if CEA, which are described by Drummond et al. (2005). The primary outcome measure may differ between interventions being investigated and therefore CEA cannot be used to make comparisons. There may also be more than one outcome of interest, which cannot be accommodated by CEA. For example, outcomes often include life extension, changes in long term quality of life and side effects. In addition, some outcomes are more important than others. For example, one treatment could extend life but cause serious side effects, and another treatment may improve quality of life but does not extend life. These two treatments cannot be compared using cost-effectiveness analysis.

This can be explained using an example. Suppose two new treatments are being considered by the NHS, one restores sight and the other restores mobility. Both treatments cost £10000 to treat one patient but the NHS can only afford to

provide funding for one of the treatments. CEA cannot be used to make such a decision as the outcomes are different and are therefore not comparable. An alternative method is to account for the change in quality of life each individual experiences after the two treatments. A quality-adjusted life year (QALY) is a measure that combines the length of time (in years) with the quality of life experienced in that time. One QALY is equivalent to one year in perfect health. A condition worth 0.2 QALYs is equivalent to an individual living in perfect health for 0.2 years. Suppose an individual is expected to experience an increase in 0.8 QALYS when their sight is restored and and an increase of 0.2 QALYs for restored mobility. The cost to restore sight can be written as $\frac{£10000}{0.8} = £12500$ per QALY and the cost to restore mobility can be written as $\frac{£10000}{0.2} = £50000$ per QALY. Using this method the treatment for restoring sight is the most cost effective. This method is called cost-utility analysis (CUA) and is used by NICE. CUA compares the cost of an intervention with the health improvement of the intervention where health improvement is measured in quality-adjusted life years (QALYs). The results of CUA are expressed as cost per QALY gained. QALYs can be calculated for any intervention and allows comparisons to be made across different interventions which usually have different primary outcomes. It is particularly useful for organisations such as NICE where decisions must be made about recommending one treatment over another. NICE uses a threshold to determine if a treatment is cost-effective, which is described on the NICE web site. If the cost of a treatment is more than $£20,000 - £30,000$ per QALY gained then it would not be considered cost-effective.

As NICE has to make decisions across different technologies and disease areas, a reference case has been defined that specifies the methods considered to be the most appropriate for the appraisal committee's purpose and consistent with the NHS objective of maximising health gain from limited resources. Assessments submitted to NICE should include an analysis of results generated using reference

case methods. The reference case is described by NICE (2008) and Claxton et al. (2005). Each element of the Health Technology appraisal has a reference case. For example, the type of economic evaluation should be cost-effectiveness analysis and the measure of health effects be the QALY.

The expected change in QALY after a treatment should be considered alongside any sources of potential bias and uncertainty. Claxton et al. (2005) and NICE (2008) discuss the importance of quantifying uncertainty associated with a health technology when assessing clinical and cost effectiveness. NICE (2008) identify three sources of uncertainty. These are structural uncertainty from the assumptions made when constructing the model, the differences between collected data, such as different costs and estimated utilities, and parameter precision(uncertainty around the mean health and cost inputs in the model). It is important to identify potential selection bias in the inputs to the model and for the model to quantify uncertainty associated with a technology. This uncertainty can be described as the probability that a different decision would be reached if the true cost effectiveness of each technology could be determined before making the decision. Methods of presenting uncertainty in cost-effectiveness are to include confidence intervals and cost-effectiveness acceptability curves. The expected mean cost and outcome should also be presented, along with the probability that the treatment is effective at the threshold of $£20,000 - £30,000$ per QALY gained and the error probability that the treatment is not cost effective. For models with few parameters analysis of the effect of the best and worst possible scenarios can be investigated. However this method is not useful is representing the combined effects of multiple sources of uncertainty when the number of parameters increase.

1.1 Outline of thesis

Conventional methods of eliciting utilities, such as the Time trade-off and Standard Gamble methods, involve questions which some respondents find difficult to answer. An alternative method is to collect discrete choice data which involves an individual choosing a preferred option from a set of health states. Utilities are required for every health state defined by a classification system. Discrete choice data is collected using a subset of health states. A model is then fitted to the data which allows the utility to be estimated for any health state defined by the classification system.

This thesis involves inference for health state utilities given discrete choice data collected using the AQL-5D classification system, which is derived from the asthma quality of life questionnaire, and also to use Bayesian Inference as a method of assessing uncertainty in parameter estimates. Chapter 2 and chapter 3 provide a background to the subject. Chapter 2 defines a QALY and classification systems, and defines methods of measuring utilities. Chapter 3 describes discrete choice data, including the process of modelling such data and a number of possible models. Both chapter 2 and 3 present a review of current relevant literature. Chapters 4, 5 and 6 present an analysis of the AQL-5D discrete choice data set. Chapter 4 uses Bayesian methods to investigate uncertainty in the model parameters. Chapter 5 presents the comparison of the models derived in chapter 4 using Bayes factors. An extension to the model in chapter 4 to include a multiplicative random effect is presented in chapter 6. An overall conclusion is reported in chapter 7.

Chapter 2

Valuing Health

2.1 Introduction

In an economic evaluation it is important to be able to make comparisons across different interventions which often have different primary outcomes. Cost-utility analysis (CUA) is a type of economic evaluation that considers quality of life, where results are expressed as cost per QALY gained. This chapter first defines the term QALY and discusses methods of measuring QALYs. The relationship between QALYs and utilities, which are both measures of preference, is also investigated. Classification systems are then defined, which are used to describe different levels of severity of health. A particular example is the AQL-5D classification system which is used to collect the data analysed in this thesis. The final section of this chapter reviews the literature in this area.

2.2 QALYs

We first consider a standard measure for health related quality of life. Often in health care the increase in both the quality and quantity of life is important. Quality-adjusted life years (QALYs) are a measure that combines the

length of time (in years) and the quality of life experienced over those years. Drummond et al. (2005) defines the QALY as a measure of health outcome that simultaneously captures gains from reduced morbidity (quality gains) and reduced mortality (quantity gains) and combines them into a single measure. Gold et al. (1996) define the QALY as a measure of health outcome which assigns to each period of time a weight, ranging from 0 to 1, corresponding to the health related quality of life during that period, where a weight of 1 corresponds to optimal health and a weight of 0 corresponds to the health state equivalent to death.

QALYs represent a length of time in years where each year is adjusted by a QALY weight which represents the quality of life experienced during that year. One QALY is equivalent to one person experiencing perfect health for a year. Alternatively one QALY could be divided between several people or years. For example, five people could each experience one year worth 0.2 QALY weight or one person could experience two years worth 0.5. Drummond et al. (2005) states that QALY weight should be based on measures of preference for a health condition, anchored on perfect health and death and be measured on an interval scale.

Brazier et al. (2007) state that for any individual if the prospect of living of living in a condition for Y years is equivalent to X years in full health where $X < Y$, then the number of QALYs experienced is X. The number of QALYs relating to a health outcome is expressed as the value given to a particular outcome multiplied by the length of time in that state. The outcome of any health intervention is uncertain and therefore the expected value of each possible outcome is weighted by its probability. The expected outcome of the treatment can be represented by the sum of the expected value of the possible individual outcomes.

The benefit of a treatment of a treatment is not the same as the outcome of the treatment as a patient may recover without treatment. The net benefit of

a treatment is the difference between the expected outcome with and without treatment.If there is no uncertainty of treatment outcomes and no changes over time the net benefit is calculated as

$$QALY_G = T_1Q_1 - T_0Q_0, \tag{2.1}$$

where T_1 is the number of years survival with treatment, T_0 is the number of tears survival without treatment, Q_1 represents the QALY weight with treatment and Q_0 represents the QALY weight without treatment.If uncertainty is introduced and changes in health can occur between time periods, the net benefit is calculated as

$$QALY_G = \sum_h \sum_t P_{1ht}Q_{ht} - \sum_h \sum_t P_{0ht}Q_{ht}, \tag{2.2}$$

where P_{1ht} and P_{0ht} represent the probabilities of an individual experiencing health state h in time period t with and without treatment respectively and Q_{ht} is the value of health state h at time t.The benefit of a treatment in the whole population is the aggregate value of the net benefits to individual patients. If the expected benefit of an average patient is obtained then the aggregate benefit of the population can be derived by multiplying the individual benefit by the number of patients expected. Therefore the QALY algorithm is a combination of the value of health states, time, probabilities and the number of patients.

2.3 Methods of measuring preferences

A health state value is a measure of preference for a particular health condition. It is measured relative to perfect health and death, which are given values of 1 and 0 respectively. Health conditions considered worse than death have negative health state values. There are several methods of measuring health state values. The three main methods used are the Time Trade-off, the Standard Gamble and the Rating Scale. These are described in sections 2.3.1, 2.3.2 and 2.3.3 respectively.

2.3.1 TIME TRADE-OFF

The Time trade-off (TTO) method was derived by Torrence et al. (1972). In this method respondents are asked to trade-off between time and quality of life. For a health state considered better than death a subject is given two alternatives and asked to choose a preferred option:

1 State j for time t followed by death.

2 Perfect health for a shortened life expectancy $x < t$ followed by death.

The time, x, is varied until the respondent is indifferent between the two alternatives. The utility of the health state, j, is then given by $\frac{x}{t}$.

For a health state, j, considered worse than death, a respondent is given two alternatives:

1 Health state j for a length of time $10 - x$ followed by perfect health for x years.

2 Immediate death.

The time, x, is varied until the respondent is indifferent between the two alternatives. The utility is given by $\frac{-x}{(10-x)}$. This allows the utilities of states worse than death to be unbounded.

2.3.2 STANDARD GAMBLE

The standard gamble (SG) method varies depending whether a health state is preferred to death or considered worse than death. Drummond et al. (2005) defines the method of measuring a utility for a health condition, j, preferred to death. A respondent is offered two alternatives:

1 The uncertain outcome of perfect health for t years with probability P and immediate death with probability $1 - P$.

2 Health state j with certainty for t years.

The probability P is varied until the subject is indifferent between the two alternatives. The utility of health state j for t years is equal to P.

McCabe et al. (2005) use a method for estimating a utility for a health condition, j, considered worse than death. The subject is given two alternatives:

1 Immediate death.

2 The uncertain outcome of perfect health with probability P and health state j with probability $1 - P$.

The utility of health state j is calculated using $U_j = -P$.

2.3.3 Rating Scale

When assigning preferences using the rating scale a subject first ranks health outcomes from the most preferred to the least preferred. The outcomes are then put on a scale where the intervals between the positions of each health state correspond to the differences in preference as perceived by the subject. Outcomes that are almost equally desirable would be placed close together and outcomes that are very different in desirability would be placed further apart. A rating scale refers to a scale of numbers, often 0-100. Category scaling and visual analogue scaling are variations of the rating scale.

Dolan (1997) and Salomon (2003) use a visual analogue scale (VAS) as a warm up exercise before providing utilities for health states using other methods. Respondents are asked to rate a predetermined set of health states on a VAS with

endpoints of 100 (best imaginable health state) and 0 (worst imaginable health state). Each state was regarded as lasting for ten years without change followed by death.

2.4 QALYs and Utilities

QALYs are used in cost-utility analysis (CUA) where the outcome of interest is cost per QALY gained. Due to the use of the word utility in CUA it is often thought that QALYs are also utilities. However, this is not necessarily true. To understand this a more formal definition of utility is defined using Bayesian decision theory, as described by O'Hagan and Forster (2004).

In health care, a utility is often called a health state utility and the health condition it is measuring a health state. Let S be an individual's health state and define $U(S)$ as the utility of state S. To measure the utility, $U(S)$, consider two other states, S_0 and S_1, where state S_1 is preferable to S and S is preferable to S_0. The utilities of S_0 and S_1 are defined as $U(S_0) = u_0$ and $U(S_1) = u_1$, where $u_0 < u_1$. Consider another state S_p where state S_1 occurs with probability p and state S_0 occurs with probability $1 - p$. The utility of S_p is defined to be

$$U(S_p) = pu_1 + (1 - p)u_0. \tag{2.3}$$

The utility $U(S_p)$ increases linearly with p. As the utility of state S is required, an individual must decide which is preferred of the two states, S and S_p. There exists a value $q \in (0, 1)$ where there is equal preference for both state S and state S_p, with $p = q$. From the definition of $U(S_p)$ the utility of state S can be written:

$$U(S) = qu_1 + (1 - q)u_0. \tag{2.4}$$

Suppose the health state of an individual depends on a decision, d. Define $S(d)$ to be the health state after making decision d and define D to be the set of all possible decisions. The optimal decision d^* will maximise utility:

$$d^* = \arg[\max_{d \epsilon D} U\{S(d)\}]. \tag{2.5}$$

The utility after making the optimal decision will be $U\{S(d^*)\}$.

Suppose the health state, S, of an individual depends on a random variable X, where the health state is now denoted by $S(X)$. Equation (2.4) can be written as

$$U\{S(X)\} = qu_1 + (1-q)u_0. \tag{2.6}$$

The random variable X has a probability distribution function $f(x)$ and a sample space $X = \{x_1, ..., x_n\}$. Suppose the value of X is known to be $X = x_i$ and $S(x_i)$ is the corresponding health state. Using Equation (2.6), the utility of $S(x_i)$ is defined to be

$$U\{S(x_i)\} = U_{x_i} = q(x_i)u_1 + (1 - q(x_i))u_0, \tag{2.7}$$

where $q(x_i)$ is the probability that health state S_1 occurs when the random variable has a value $X = x_i$. The state S will change to state $S(x_i)$ if $X = x_i$, which has a probability of $f(x_i)$ of happening. The health state S in Equation (2.4) is defined to have an equal preference to health state S_p when $p = q$. Similarly, when the random variable X is involved, health state S can be considered to have equal preference to health state S_p where S_1 occurs with probability

$$p = \sum_{i=1}^{n} f(x_i)q(x_i) \tag{2.8}$$

and S_0 occurs with probability

$$1 - p = \sum_{i=1}^{n} f(x_i)(1 - q(x_i)) \tag{2.9}$$

The utility of S can therefore be written as:

$$\begin{aligned} U(S) &= U(S_p) = pu_1 + (1-p)u_0 \\ &= \sum_{i=1}^{n} f(x_i)\{q(x_i)u_1 + (1 - q(x_i))u_0\} \\ &= \sum_{i=1}^{n} f(x_i)U(S(x_i)) \\ &= E\{U(S(X))\} \end{aligned} \tag{2.10}$$

Therefore if the state S is a function of an unknown variable X the utility is calculated by taking the expectation of $U\{S(X)\}$. This result is known as the law of expected utility.

The state following a decision can be written as a function of both the decision d and the unknown variables X: $S(d, X)$. The optimum decision d^* will be the one that maximises utility:

$$d^* = \arg[\max_{d \epsilon D} E[U\{S(d, X)\}]] \tag{2.11}$$

Choosing the decision that maximises expected utility is known as Bayes' rule.

A utility is measured using a method involving uncertainty. Health state values can only be considered utilities if they are elicited using a method with uncertainty. The only standard method to measure health state values that involves uncertainty is the Standard Gamble. Health state values derived using a method with no uncertainty are not necessarily utilities. The relationship between health

state values and utilities depends on risk attitude. If an individual is risk neutral then they are equal. However, if an individual is risk averse then their utility is expected to be greater than their value and if an individual is risk seeking their utility is expected to be less than their value.

Drummond et al. (2005) suggest that only QALYs derived using preferences measured using the Standard Gamble can be considered utilities. However, they must follow certain assumptions, which are also presented by Brazier and Tsuchiya (2010). These are that the two attributes of quality and quantity must be mutually utility independent, the proportion of remaining life that would be used as a trade-off for a specified quality improvement is independent of the amount of remaining life and the single attribute utility function for additional healthy life years must be linear with time. In this thesis the term utility will refer to any health state preference, whether or not it satisfies the definition of utility described previously.

2.5 Classification systems

Given the number and variety of illnesses treated by the NHS, it would not be possible to assess utility for every health condition of interest. A finite set of heath conditions needs to be defined so that utilities can be elicited for every possible condition. In general, the approach used is first to consider a small list of attributes of a health condition, e.g. pain, mobility. For each attribute, various levels of severity are defined, e.g. no pain, moderate pain, severe pain. The combination of levels for each of the attributes then defines a health state. The set of attributes and levels that defines a health state is called a classification system. Drummond et al. (2005) describes some examples of health state classification systems. The EQ-5D, SF-6D and the Health Utilities Index all describe general health and can be used for any illness. These are described in Section 2.5.1,

2.5.2 and 2.5.3. Some classification systems are used only for a specific health condition. An example is the AQLQ which is used for asthma and is described in Section 2.5.4.

2.5.1 EQ-5D

The EQ-5D (or EuroQol) system, as used by Dolan (1997) and described by EuroQol (1990) and Brooks and Group (1996), has five attributes: mobility, self-care, usual activity, pain/discomfort and anxiety/depression. Each attribute can take one of three levels: no problem (level 1), some problems (level 2) and major problems (level 3). After including two additional heath states, dead and unconscious, the system describes 245 health states. The health state described when all attributes are at level 1 is perfect health. The worst possible health state is indicated when all attributes are at level 3. An example of a health state is 12233. This describes the health state where there are no problems walking about, some problems washing and dressing, some problems performing usual activities (e.g.work), extreme pain or discomfort and extreme anxiety or depression.

2.5.2 SF-6D

The SF-6D, as discussed by McCabe et al. (2006) and Brazier et al. (2002), is a classification system composed of six attributes: physical functioning, role limitations, social functioning, pain, mental health and vitality. Each attribute has between four and six levels, defining 18000 unique health states.

2.5.3 Health Utilities Index

The Health Utilities Index (HUI), as reviewed by Feeney et al. (1995) and Feeney et al. (2002) consists of two systems, the HUI2 and the HUI3. The HUI3 has eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and

pain. Each attribute has either 5 or 6 levels. The HUI2 has some additional attributes that are not used in HUI3 that may be used in specific studies. HUI2 has the attributes sensation, mobility, emotion, cognition, self-care and pain. There is also an optional attribute fertility. Each attribute has either 3, 4 or 5 levels.

2.5.4 AQLQ

The asthma quality of life questionnaire (AQLQ) is a 32 item questionnaire developed to measure the functional impairments that are most important for adults with asthma. The 32 items have 7 levels each and the items cover the domains: symptoms (12 items), activity limitations (11 items), emotional function (5 items) and environmental stimuli (4 items). Yang et al. (2006) describe how this has been reduced to a 5 attribute classification system called AQL-5D. The attributes are: concern about asthma, short of breath, weather and pollution stimuli, sleep impact and activity limitations. Each attribute has 5 levels of severity with level 1 indicating no problem and level 5 extreme problems, thus defining 3125 health states. The AQL-5D has been used to collect the data analysed in this thesis. Table 2.1 lists the attributes and the levels of the AQL-5D. Each attribute has the same five levels.

Attributes		Levels	
1	Feel concerned about having Asthma	1	None of the time
2	Feel short of breath as a result of asthma	2	A little or hardly any of the time
3	Experience asthma as a result of air pollution	3	some of the time
4	Asthma interferes with sleep	4	Most of the time
5	Activities have been limited	5	All of the time

Table 2.1: Attributes and levels of AQL-5D classification system

2.5.5 Comparison of generic and condition-specific classification systems

Health state utilities calculated using classification systems that describe general health are regarded as being applicable to any health condition and patient group. NICE prefers the use of the EQ-5D to calculate QALYs for use in the economic evaluation of all interventions being investigated. It is thought that to achieve comparability between utilities a common classification system must be used. A reason for NICE preferring the use of the EQ-5D is that it ensures that different patient groups are being assessed against the same attribute and can be compared.

For some conditions, utilities derived using generic classification systems are not appropriate. Espallargues et al. (2005), Barton et al. (2004), Walters et al. (1999) and Haywood et al. (2008) compare classification systems for the estimation of utilities for specific conditions. Espallargues et al. (2005) estimated health utilities for age related macular degeneration (ARMD) associated with visual impairments using the EQ-5D, the SF-6D and the HUI3. They conclude that the EQ-5D and the SF-6D were not relevant to the condition but the HUI3 was more appropriate due to the inclusion of an attribute for vision. Barton et al. (2004) also compared the EQ-5D, SF-6D and the HUI3. Utilities were derived for hearing impaired adults both before and after being given a hearing aid. It was found that the mean change in utility was significantly higher when using the HUI3 system than for either the EQ-5D or the SF-6D. The estimated cost-effectiveness would therefore depend on the classification system used. Walters et al. (1999) compare the EQ-5D, SF-6D, SF-MPQ and the FAI classification systems for estimating utilities for venous leg ulcers. The SF-MPQ, which defines the intensity of pain, is described by Melzack (1987), and the FAI, originally designed for stroke patients, is described further by Holbrook and Skilbeck (1983). The classification system recommended by Walters et al. (1999) depends on the length of follow-up

and sample size. Haywood et al. (2008) compares the EQ-5D with two condition specific classification systems for deriving utilities for women with urinary incontinence. The two condition-specific systems are the symptom severity index (SSI), described by Black et al. (1996), and the urinary incontinence quality of life instrument (I-QoL), described by Wagner et al. (1996). The I-QoL is the only classification system recommended for use in this case.

As the use of general classification systems to calculate utilities is not necessarily an accurate method for some conditions, the use of condition-specific measures (CSMs) is often preferred in economic evaluation. However, if the CSMs are not preference-based measures they cannot be used to calculate QALYs. In these cases the CSM value is mapped onto values derived from a general classification system using regression techniques. This is designed to be a more accurate method of estimating utilities for specific conditions than using a generic classification system. Tsuchiya et al. (2002) describes several models for mapping such values. Brazier et al. (2008) reviews 28 studies which use functions to map between non-preference based condition-specific outcomes and generic preference-based measures. An example of a study reviewed is Longworth et al. (2005), which investigates the relationship between angina-specific outcomes and utilities derived for angina patients using the EQ-5D. Brazier et al. (2008) found that the goodness of fit and prediction of the mapping functions was variable and so it is not possible to generalise across classification systems. The models were also limited in their ability to predict means for subgroups.

An alternative to mapping is to estimate utilities using a condition-specific classification system. Brazier and Tsuchiya (2010) investigate the problems associated with comparing utilities across different interventions where each uses a different classification systems. It is suggested that utilities can be compared when derived from different classification systems provided that they meet some con-

ditions. These conditions are that the same method of eliciting utilities is used, perfect health and death are always used as anchors and the respondents used are selected from the same population group. However, there are also several issues that may affect achieving comparability, which are as follows:

1 A condition-specific classification system does not allow for important side-effects which would be covered using a generic classification system.

2 Patients and non-patients might have different interpretations of attributes.

3 The upper anchor used is regarded as the best health state as defined by the classification system and therefore different classification systems have a different definition for the best health state.

4 Respondents tend to focus on problems described in the health state they are valuing and this results in exaggerating the importance of the problems associated with the condition being valued compared to other conditions.

5 The achievement of comparability between classification systems requires the assumption that the impact of different attributes on utilities are additive even if the attributes are not included in the classification system. If an intervention changes an attribute level in a classification system the estimated change in utility may be incorrect due to interactions between attributes included in the classification system and attributes not included. Brazier and Tsuchiya (2010) suggest that such interactions may be larger for condition-specific classification systems since they focus on a narrower range of health states.

Brazier and Tsuchiya (2010) recommend the use of condition-specific classification systems as a supplement to generic classification systems rather than an alternative.

2.6 Inference for health state utilities given elicited preferences

For the remainder of this thesis we refer to health states to mean a health state as defined by a particular classification system. In an economic evaluation of a treatment of a health condition, estimated mean utilities are required for each health state of the condition. These mean utilities are then used to calculate QALYs. To achieve this aim, a survey is conducted to elicit utilities for a subset of the health states in the classifications system from a sample of the population. A model is then fitted to the data which can be used to estimate the utility of any health state defined for that condition.

2.6.1 Modelling multi-attribute health state utilities

Basu et al. (2009) develop a model to predict the utilities of health states defined by several attributes, using utilities of the component single attribute health states. Three methods are reviewed to estimate these multi-attribute utilities; these are the additive model, the multiplicative model and the minimum model. To explain these first let $SS1$ and $SS2$ to be two single attribute health states and let JS be the corresponding two attribute health state where $JS = (SS1, SS2)$. Define $U(SS1)$ to be the utility of health state $SS1$ and $U(SS2)$ to be the utility of health state $SS2$. The expected utility of the health state JS using each of the three models is shown as follows.

Additive Model

In the additive model the two utilities are added together and the utility of perfect health, 1, is subtracted.

$$E[U(JS)] = U(SS1) + U(SS2) - 1. \tag{2.12}$$

For example, if $U(SS1) = 0.8$ and $U(SS2) = 0.7$ then $E[U(JS)] = 0.5$. This shows the utility decreases if both attributes are present in a health state.

Multiplicative Model

The estimated utility using the multiplicative model is

$$E[U(JS)] = U(SS1) \times U(SS2). \tag{2.13}$$

For example, if $U(SS1) = 0.8$ and $U(SS2) = 0.7$ then $E[U(JS)] = 0.56$.

Minimum Model

$$E[U(JS)] = min\{U(SS1), U(SS2)\}. \tag{2.14}$$

For example, if $U(SS1) = 0.8$ and $U(SS2) = 0.7$ then $E[U(JS)] = 0.7$.

Basu et al. (2009) reference the work in Keeney and Raiffa (1976) and Keeney and Raiffa (1993) about a model using both additive and multiplicative interaction terms. Letting $U(JS)$, $SS1$, $SS2$, $U(SS1)$ and $U(SS2)$ be as previously defined, the above model is

$$U(JS) = k_1.U(SS1) + k_2.U(SS2) = k.k_1.k_2.U(SS1).U(SS2). \tag{2.15}$$

The weights k, k_1 and k_2 can be estimated from elicited utilities of a multi-attribute health state or directly from respondents, although it is unclear how the second method is carried out. Basu et al. (2009) use the model in equation (2.15) as a starting point to suggest a model which combines elements of the additive, multiplicative and minimum models. The model is presented in terms of utility loss rather than utility. A utility of U for any health state indicates a

loss of $1-U$ utility units from perfect health, which has a utility of 1. Define $L(SS1)$ be the loss of utility for health state $SS1$ and $L(SS2)$ be the loss of utility for health state $SS2$. The expected loss of utility for health state JS is

$$E[L(JS)] = \alpha_0+\alpha_1 max\{L(SS1), L(SS2)\}+\alpha_2 min\{L(SS1), L(SS2)\}+\alpha_3 L(SS1)L(SS2), \tag{2.16}$$

where the parameters α_0, α_1, α_2 and α_3 are weights. The additive, multiplicative and minimum models are special cases of this model depending on the values given to the parameters. The expected utility is then given by $E[U(JS)] = 1 - E[L(JS)]$.

Dale et al. (2008) investigate which of the additive, multiplicative and minimum models, are the most appropriate to estimate multi-attribute health states for prostate cancer. It was concluded that all three models gave biased predictions. If it is not possible to elicit utilities directly for multi-attribute states then the minimum model is recommended. Stewart et al. (2005) also investigate the use of both the additive and multiplicative models in estimating multi-attribute health states for prostate cancer in order to develop a decision model of outcomes for prostate cancer treatment. Plante et al. (1987) construct a decision tree to determine the preferred treatment option for stage 3 squamous cell carcinoma. The decision tree is based on quality adjusted weeks of survival. The life expectancy of the population used in the the trial is estimated in weeks using life tables and is adjusted by QALYs to estimate the quality adjusted weeks of survival. QALYs were assigned to four treatment options based on categorical scaling and published values for other major diseases. In cases where more than one treatment is used, QALYs are estimated using the multiplicative model.

Sutherland et al. (1982) investigate the attitudes to duration of survival in different hypothetical health states. Preferences were assessed using two methods:

a preference questionnaire and a certainty equivalence method. In the questionnaire respondents were asked if a defined period of time in a certain health state was preferred to death. The certainty equivalence method involves respondents choosing between a period of guaranteed survival in a given health state and a gamble between surviving for a longer period of time in the same health state or dying immediately. It was found that attitudes toward duration of survival depended on the quality of the health state.

2.6.2 Modelling health state classification system data

This section reviews the literature for each of the main classification systems and the estimation of multi attribute utilities using single attribute utilities.

The EQ-5D is first described by EuroQol (1990). Some of the first studies conducted using this system are presented by Nord (1991) and Brooks et al. (1991) which fit models to VAS scores collected for a subset of EQ-5D health states in order to predict utilities for any health state in the system. Dolan (1997) present a good example of a model used to analyse an EQ-5D data. From the 243 possible health states that the EQ-5D system defines, a subset of 42 health states were selected. The health states included as many combinations of levels across the attributes as possible. Each respondent was asked to value 13 of the 42 health states using the Time Trade-off method. The model for the utility is:

$$U_{ij} = \delta + \boldsymbol{h}(\boldsymbol{x}_{ij})^T \boldsymbol{\theta} + \varepsilon_{ij}, \qquad (2.17)$$

where U_{ij} is the utility of health state x_{ij} for individual i, δ is the intercept, $\boldsymbol{x}_{ij}^T$ is the vector of explanatory variables defining the health state, $\boldsymbol{h}$ is a known vector of functions of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is the vector of corresponding unknown parameters. In addition to variables representing each attribute level, the vector $\boldsymbol{x}$ includes

variables representing the change in level of an attribute, interactions between attributes, and variables that count the number of times a health state contains an attribute at either level 1 or level 3.

Dolan et al. (1996a) also reports the results using the same study as in Dolan (1997) and investigate differences between subgroups in the population. The results showed that valuations for severe health states appeared to be affected by age and sex and it is suggested that small samples used in other studies may be concealing real differences that exist between groups in the population.

Dolan and Roberts (2002) also use the same data set as used in Dolan (1997) to fit a model estimating the difference in utility between the worst health state defined by the EQ-5D and all other health states in the system. The differences are explained in terms of the change in level of each attribute. It was found that the new model is more accurate at predicting the utilities of states that have been directly observed than the original model in Dolan (1997).

Hoeymans et al. (2005) fit a similar model to Dolan (1997) using the EQ-6D classification system, which is the EQ-5D with an additional attribute called cognitive functioning. A model was constructed for each attribute separately where the values were explained by variables such as age and sex. It was also assumed that there was a higher correlation between certain attributes.

2.6.3 Bayesian Models

Kharroubi et al. (2005) and Kharroubi et al. (2007) developed a nonparametric Bayesian model for estimating utilities of health states in the SF-6D descriptive system. A subset of 249 of the 18000 possible states were used in the study. A total of 836 respondents were asked to value 6 of these health states. Preferences

were elicited using the standard gamble method described in Section 2.4. The proposed model is:

$$U_{ij} = 1 - \gamma_i(1 - g(\boldsymbol{x}_{ij})) + \varepsilon_{ij} \tag{2.18}$$

where U_{ij} is the utility that is elicited from respondent i for health state $\boldsymbol{x}_{ij}$, $g(\boldsymbol{x}_{ij})$ is the population mean utility for health state x_{ij}, γ_i is a random effect for respondent i and ε_{ij} is the independent $N(0, \sigma^2)$ error.

When the health state is perfect health, $g(\boldsymbol{x}_{ij}) = 1$. The random effect term is multiplicative, allowing the elicited utilities for poorer health states to have greater variability than utilities for better health states and accounts for respondents attitudes to poorer health states. When $\gamma_i = 1$ the respondent's utilities are in agreement with the population average and $E(U_{ij}) = g(\boldsymbol{x}_{ij})$. If $\gamma_i < 1$ the respondent is less worried by the prospect of poor health and would have utilities greater than the population mean. If $\gamma_i > 1$ the respondent is more worried about poor health than the population average and would tend to have utilities greater than the population mean utility. The respondents with the largest value of γ_i are more likely to value states as worse than death.

The utility function was modelled using Bayesian hierarchical modelling. There are L different health states defined by the classification system and therefore L unknown utilities $g(\boldsymbol{x}_{ij})$. The prior distribution specifies that the utilities have an L-dimensional multivariate normal distribution. The Bayesian hierarchical prior structure incorporates two prior beliefs about the function g. The first prior belief is that poorer health states should have lower utility and that g should be a decreasing function of the dimensions of the health state. This is represented by

$$E\{g(\boldsymbol{x}_{ij}|\theta\} = h(\boldsymbol{x}_{ij})^T\theta$$

where $h(.) = (h_1(.), h_2(.), .., h_p(.))^T$ is a vector of p functions of the health state

x and $\theta = (\theta_1, \theta_2, ..., \theta_p)$. The choice for $h(x)$ is $h(x) = (1, x_1, x_2, x_3, x_4, x_5, x_6)^T$ representing a prior belief that utility will be approximately linear and additive in the different dimensions, where $x_1, x_2, ..., x_6$ are variables representing the levels of each of the six attributes in the SF-6D classification system.

The second prior belief about function g is that if $\boldsymbol{x}$ and $\boldsymbol{x}'$ describe similar health states they should have similar utilities. This implies that there is expected to be high correlation between $g(\boldsymbol{x})$ and $g(\boldsymbol{x}')$. The correlation should decrease as the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ increases. The covariance is defined as

$$cov\{g(\boldsymbol{x}), g(\boldsymbol{x}')|\sigma^2\} = \sigma^2 \exp\{-\sum b_d(x_d - x'_d)^2\} \tag{2.19}$$

where, x_d is the level of attribute d for health state $\boldsymbol{x}$, x'_d is the level of attribute d for health state $\boldsymbol{x}'$, and b_d is a roughness parameter for attribute d. For this model, $b_d = \dfrac{2.5}{(l_d - 1)^2}$, where l_d is the number of levels in attribute d.

The prior distribution for the random respondent effect, α_i is assumed to be an independently log Normal distribution, $\ell N(0, \tau^2)$. The parameters υ^2, τ^2, γ, θ and σ^2 are all given weak prior distributions, written

$$p(\upsilon^2, \tau^2, \gamma, \theta, \sigma^2) \propto \upsilon^{-2}\sigma^{-1}\tau^{-2}.$$

An MCMC was run with 3000 iterations to compute posterior inferences about the population mean utility. It was concluded that the model used represents certain important characteristics of the data more accurately than previously proposed models. These characteristics include individual response effects, repeated measurements from each individual, the skew distribution of individual valuations of a given health state, and the nonparametric relationship between health state and utility.

2.6.4 Measuring utility of death and Utility Scale

Utilities are defined on a scale relative to the utilities of perfect health and death. Perfect health usually has a utility of 1 and death has a utility of 0. Health states considered worse than death have negative utilities.

Dolan (1997) consider the utility scale by discussing how to interpret the intercept, δ, in Equation (2.17). When the dummy variables in $\boldsymbol{x}_{ij}$ are all zero, the estimate of $h(\boldsymbol{x}_{ij})^T\boldsymbol{\theta}$ is for the health state perfect health. The utility U_{ij} is defined to be 1 when the health state is perfect health. The intercept, δ, is therefore interpreted as the estimated value for 1 minus the estimated mean utility for perfect health. The mean utility for perfect health is therefore $1-\delta$. Estimated utilities are then divided by $1-\delta$ to ensure that utilities are on a scale where the utility of perfect health is 1. Dolan (1997) suggested that δ could alternatively represent the additional change in utility for any health state with at least one attribute at level 2, and therefore represent any move away from perfect health. When predicted and actual values were compared, the method in which δ was treated this way performed much better than when all estimates were divided by $1-\delta$. The model possibly should not have included the intercept, δ, ensuring that full health always has an estimated utility of 1. The estimated value of the parameter δ suggests that more interaction terms should be used in the model.

2.6.5 Comparing methods of preference

Spencer (2003) investigates whether the inferences about people's preferences towards health states vary if the TTO procedure used to elicit preferences is varied. Respondents were asked to answer two sets of TTO questions regarding the health states in a EQ-5D classification system. Some levels were changed to ensure that health states considered worse than death were not included in the study.

Respondents were asked four conventional TTO questions, involving health states from the EQ-5D classification system: 12221, 21211, 21222 and 22232. These health states are referred to by the letters W, X, Y and Z respectively. The conventional TTO method is the method described in Section 2 for health states considered better than death, where the length of time, t, of being in a particular health state is 10 years.

When using the unconventional TTO method, respondents are asked to imagine living in a given health state, S_1, for two years followed by death, or prolonging life in a lower quality of life, health state S_2. A respondent is given two alternatives:

1 Health state S_1 for two years, followed by death.

2 Health state S_2 for a time x_1 where $x_1 > 2$.

The time x_1 is increased until the respondent is indifferent between the two alternatives. The utility of health state S_1 is elicited using conventional TTO methods and given a utility of $U(S_1) = \frac{x}{10}$. The utility of health state S_2 is then given by

$$U(S_2) = \frac{2 \times x}{10 \times x_1}. \tag{2.20}$$

Each respondent was asked two unconventional questions. The first involved eliciting a utility for health state Z by comparing it to health state Y, where $S_1 = Y$ and $S_2 = Z$. The second unconventional TTO question involved comparing health state Y with health state X, where $S_1 = X$ and $S_2 = Y$.

Each of the two health states Y and Z has a conventional and an unconventional TTO value. Define the conventional TTO values of health states Y and Z to be U_y and U_z, and define the unconventional TTO values of health states

Y and Z to be V_y and V_z. The null hypothesis tested is that

$$U_y = V_y$$

and

$$U_z = V_z.$$

A Wilcoxon signed rank test was used to test the null hypothesis that the conventional and unconventional elicited utilities are equal. The alternative hypothesis was that there is a systematic difference between responses.

The test showed that the unconventional values for Y were significantly lower than the conventional values. The differences between the unconventional and conventional values for Z were found to be not statistically significant. As only one test detected a statistical difference the conclusion of the study was that there was no evidence to reject procedural invariance. As only one of the two tests detected a statistical difference, a more appropriate conclusion may have been to reject the null hypothesis.

Salomon and Murray (2004) present a new methodological approach that allows estimation of a set of health states through multiple measurement techniques. The techniques used were the visual analogue scale, standard gamble, time trade-off and person trade-off. The health state classification system used included the 5 attributes used in the EQ-5D system and an additional attribute, cognition. Each attribute in the system has five levels of severity. Twelve health states were selected from the classification system and respondents were asked to state preferences using the four methods. Each measurement technique produced responses from individuals for health states on a scale particular to the method used. Responses from each of the four methods are transformed onto a utility scale between 0 and 1 where death has a utility of 0 and perfect health has a

utility of 1.

A model was fitted that related the utility to the measurement type used. Maximum likelihood methods were used to estimate the parameters. The re-scaled responses from each method were assumed to have a truncated normal distribution, constrained between 0 and 1 with parameters specific to state and method. As the variance between responses varied across methods and was strongly related to the mean value of each state, the variance was modelled as a linear function of the mean allowing the slope and intercept to differ for each measurement method.

Dolan et al. (1996b) discuss the issue that utilities elicited using the SG method and those elicited using the TTO do not give the same value for a given respondent. The study compares four sets of utilities for a set of EQ-5D health states. The two methods of eliciting utilities were the SG and TTO. Each method had two variants, one where specially designed boards and cards were used as an aid to decision making by respondents and the other involved the use of a self-completed book. Comparison of methods has also been reported by Torrence (1976), where the SG, TTO and category scaling methods were analysed for their feasibility, reliability, validity and compatibility. Read et al. (1984) compare the SG, TTO and category scaling for assessing preferences among hypothetical outcomes of coronary artery bypass surgery.

Several articles compare classification systems for measuring health related quality of life. McDonough et al. (2005) compare utilities elicited using the EQ-5D, HUI and SF-36 for people diagnosed with intervertebral disc herniation, spinal stenosis or degenerative spondylolisthesis. Utilities were elicited using the TTO method for the EQ-5D system, using SG and VAS for the HUI, and using SG for the SF-36. Summary statistics were estimated for each classification system and comparisons between distributions for each pair of systems were made using the

Wilcoxon signed rank tests. Stavem et al. (2005) compare utilities derived using the classification systems 15D (www.15d-instrument.net) with those derived using the EQ-5D and SF-6D for patients with HIV/AIDS. Brazier et al. (1993) compare the EQ-5D with the SF-36.

2.7 Summary

This chapter reviewed QALYs which are used as a measure of valuing health care where utilities are used as a measure of preference for each health state. Methods of measuring utilities such as the Standard Gamble and Time Trade-off are defined. A health state classification system is often used to define a finite set of health states. These can be disease-specific or generic. There is some debate in the literature about which method is most appropriate as estimated utilities are not always consistent. The preferred method appears to depend on the condition being investigated. Models fitted in the literature involve estimating utilities using either classification system or estimating multi-attribute utilities from single attribute utilities.

Chapter 3

Discrete Choice data

3.1 Introduction

Conventional methods of eliciting utilities such as the TTO and SG were reviewed in chapter 2. These methods involve questions that some respondents find difficult to answer. The Standard Gamble needs the respondent to understand probability and in the Time Trade-off method an individual might find it difficult to imagine living for a finite number of years. Both methods also require a person being certain of their indifference between the two alternatives for a given P or t. There are also concerns, discussed in Brazier et al. (2006), that observed utilities are related to factors other than the respondent's preference for the health states. These factors include risk aversion for the Standard Gamble, and time preference and aversion to losses for the Time Trade-off. This chapter reviews discrete choice experiments, which are used to collect discrete choice data and is regarded as an easier method of collected preferences than the standard methods reviewed. The process of modelling such data is explained along with properties and identifiability issues often associated with these types of models. Three models are defined: the logit, probit and mixed logit. Finally a review of relevant literature is undertaken.

3.2 Discrete choice experiments

A discrete choice experiment (DCE) involves respondents choosing which of a list of alternatives they prefer. First, a set of alternatives is defined. Then a subset of alternatives is selected and assigned into groups. Each respondent is asked which alternative in a group they prefer. This is carried out for several groups. This type of data is called discrete choice data.

A discrete choice experiment can involve respondents choosing which of two alternatives they prefer. Such assessments are called pairwise comparisons and are carried out for several pairs of alternatives.

3.3 Modelling discrete choices

We wish to obtain the population mean utility for every alternative in a defined set of alternatives. The population mean utility can be inferred using certain modelling assumptions, which are discussed in this section.

Let $\boldsymbol{x}$ be a vector of dummy variables that defines an alternative. For example, if a set of possible alternatives is defined using 3 attributes where each attribute has 3 levels of severity, $\boldsymbol{x}$ would be a vector of 6 variables, $\boldsymbol{x} = (x_1, x_2, ..., x_6)$. For $a = 1, 2, 3$ and $b = 1, 2$ each element of $\boldsymbol{x}$ is defined as

$$x_{2(a-1)+b} = \begin{cases} 1 & \text{if attribute } a \text{ is at level } b+1 \text{ or higher} \\ 0 & \text{if attribute } a \text{ is at level 1} \end{cases} \tag{3.1}$$

Suppose an individual i considers the alternatives in the set $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iJ}\}$. Define U_{ij} to be the utility individual i has for alternatives $\boldsymbol{x}_{ij}$. The relationship

between U_{ij} and $\boldsymbol{x}_{ij}$, which is used in McFadden (1974), can be expressed as

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \quad j = 1, ..., J \quad , \tag{3.2}$$

where $g(\boldsymbol{x}_{ij})$ is a function of $\boldsymbol{x}_{ij}$ with unknown parameters and represents the population mean utility for alternative $\boldsymbol{x}_{ij}$. In this thesis, we consider population mean utility functions of the form $g(\boldsymbol{x}_{ij}) = 1 - \boldsymbol{\theta}^T \boldsymbol{x}_{ij}$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. The error term, ε_{ij}, represents the individual's variation in preference from the population mean utility.

In a discrete choice experiment an individual i selects the preferred alternative in set B. The probability that individual i chooses alternative $\boldsymbol{x}_{ij}$ is equivalent to the probability that the utility individual i has for alternative $\boldsymbol{x}_{ij}$ is greater than the utility individual i has for all the other alternatives in set B. Define $P_B(\boldsymbol{x}_{ij})$ to be the probability that individual i will choose alternative $\boldsymbol{x}_{ij}$ from the set B. The probability can be written as

$$P_B(\boldsymbol{x}_{ij}) = P[g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} \quad \text{for all} \quad k \neq j] \tag{3.3}$$

Given the value of $g(\boldsymbol{x}_{ij}), \forall j$, the probability $P_B(\boldsymbol{x}_{ij})$ will depend on the distribution of the error terms $\varepsilon_{i1}, \varepsilon_{i2}, ...\varepsilon_{iK}$. The distribution assumed determines the model that is used, though choosing the distribution is not straightforward. The errors could be assumed to have a normal distribution, and for discrete choice data this is called a probit model. The errors are also often assumed to have a type 1 extreme value distribution, which is a logit model. Logit models and probit models are discussed further in Sections 3.5 and 3.6.

3.4 Identifiability in discrete choice models

The likelihoods for models such as the logit and probit are derived from equation (3.3). As we do not observe U_{ij} directly, there are limitations on how the population mean utility and the error can be specified as not all parameters are identifiable. These methods are described by Train (2003).

3.4.1 Identifiability of Additive Constants

The utility individual i has for alternative $\boldsymbol{x}_{ij}$ is given by equation (3.2). Utility is measured relative to the utilities of two fixed alternatives. Therefore, the utility scale is determined by the utilities given to the two fixed alternatives. Suppose for individual i, $U_{i1} < U_{i2} < ... < U_{iJ}$. If each utility, U_{ij}, is transformed onto a new scale by adding a constant δ to each utility, then $U_{i1}+\delta < U_{i2}+\delta < ... < U_{iJ}+\delta$. On the transformed scale, the ordering of the utilities is the same and individual i would make exactly the same choices in a discrete choice experiment. We write individual i's utilities on this transformed scale as

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \delta + \varepsilon_{ij} \quad j = 1, ...J. \tag{3.4}$$

The probability, $P_B(\boldsymbol{x}_{iJ})$, of choosing alternative $\boldsymbol{x}_{ij}$ from the set of alternatives $B = \{\boldsymbol{x}_{i1}, ...\boldsymbol{x}_{ik}\}$, for the utility defined in equation (3.4) is

$$\begin{aligned} P_B(\boldsymbol{x}_{ij}) &= P[g(\boldsymbol{x}_{ik}) + \delta + \varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \delta + \varepsilon_{ij} \quad \text{for all} \quad k \neq j] \\ &= P[g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} \quad \text{for all} \quad k \neq j], \end{aligned} \tag{3.5}$$

which is the same as the probability without the constant, δ, in equation (3.3). The constant in equation (3.4) cannot be estimated as any value will give the same probability, $P_B(\boldsymbol{x}_{ij})$.

Constants can be included in the model if they are specified in a way that creates differences in utility over alternatives. Train (2003) states that to include additive constants in a model one constant must be normalised to zero and all other constants estimated relative to the normalised constant. Train (2003) explains normalising constants by describing an example of including a covariate in the model. Consider the case where the utility an individual has for an alternative varies over age. Define Y_i to be a variable representing the age (in years) of individual i. The utilities, U_{i1} and U_{i2}, that individual i has for alternatives $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ are

$$\begin{aligned} U_{i1} &= g(\boldsymbol{x}_{i1}) + Y_i\beta_1^1 + \varepsilon_{i1} \\ U_{i2} &= g(\boldsymbol{x}_{i2}) + Y_i\beta_2^1 + \varepsilon_{i2}, \end{aligned} \tag{3.6}$$

where β_1^1 and β_2^1 capture the effects that a change in age has on the utility of alternatives $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$. Utility is assumed to increase with age, so $\beta_1^1 > 0$ and $\beta_2^1 > 0$, and age is assumed to have a different effect on the utility of each alternative, so that $\beta_1^1 \neq \beta_2^1$. The probability $P_B(\boldsymbol{x}_{i1})$ of individual i preferring health state $\boldsymbol{x}_{i1}$ to health state $\boldsymbol{x}_{i2}$ is

$$\begin{aligned} P_B(\boldsymbol{x}_{i1}) &= P[g(\boldsymbol{x}_{i2}) + Y_i\beta_2^1 + \varepsilon_{i2} < g(\boldsymbol{x}_{i1}) + Y_i\beta_1^1 + \varepsilon_{i1}] \\ &= P[\varepsilon_{i1} - \varepsilon_{i2} > g(\boldsymbol{x}_{i2}) - g(\boldsymbol{x}_{i1}) + Y_i(\beta_2^1 - \beta_1^1)], \end{aligned} \tag{3.7}$$

It is not possible to identify values for both β_1^1 and β_2^1; the likelihood will be unchanged for any value of β_1^1 and β_2^1. One parameter is therefore normalised to zero, by subtracting $Y_i\beta_1^1$ from each of the utilities in the equation 3.6. The

utilities are now

$$\begin{aligned} U_{i1} &= g(\boldsymbol{x}_{i1}) + \varepsilon_{i1} \\ U_{i2} &= g(\boldsymbol{x}_{i2}) + Y_i\boldsymbol{\beta}_2 + \varepsilon_{i2}, \end{aligned} \tag{3.8}$$

where $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^1 - \boldsymbol{\beta}_1^1$ and is interpreted as the differential effect of age on the utility of the two alternatives $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$. The value of $\boldsymbol{\beta}_2$ can be positive or negative.

It would not be possible to include a socio-demographic variable such as $Y_i\beta_1^1$ for every possible alternative in a discrete choice experiment, as data is usually only collected for a subset of alternatives. To estimate an age parameter for every alternative, a model could include interactions between the age variables and dummy variable defining the alternative. Train (2003) states that when sociodemographic variables are interacted with the attributes of the alternatives, coefficients do not need to be normalised. The sociodemographic variables affect the differences in utility through their interaction with the attributes of the alternatives. This method would not always be appropriate as not all data sets would be designed to include interactions. A more suitable method could be to assume age has an equal affect on every alternative or to include it as the mean of a multiplicative random effect.

3.4.2 Identifiability of Error variances

Discrete choice data is modelled by assuming a distribution for the errors, ε_{ij}. It is important to consider properties that errors can have as these determine the utility scale. Multiplying a utility by a positive constant will change the variance of the error distribution. Utilities are defined on a fixed scale and multiplicative

constants can be used to normalise the utility scale by fixing the variance. This section discusses how to normalise the utility scale for independent and identical errors, heteroscedastic errors and correlated errors. These methods are reviewed by Train (2003).

Independent Errors

Suppose the utility individual i has for alternative $\boldsymbol{x}_{ij}$ is described by equation (3.2), where the error term, ε_{ij}, is assumed to be independently and identically distributed with variance $Var(\varepsilon_{ij}) = \sigma^2$.

Suppose for individual i, $U_{i1} < U_{i2} < .... < U_{iJ}$. If each utility $U_{ij}, j = 1, ..., J$ is transformed onto a different scale by multiplying by a positive constant, λ, then $\lambda U_{i1} < \lambda U_{i2} < ... < \lambda U_{iJ}$. The utilities on the transformed scale are written as

$$\lambda U_{ij} = \lambda g(\boldsymbol{x}_{ij}) + \lambda \varepsilon_{ij}, \quad j = 1, ..., J, \tag{3.9}$$

where the error, $\lambda\varepsilon_{ij}$, has variance $Var(\lambda\varepsilon_{ij}) = \lambda^2 Var(\varepsilon_{ij}) = \lambda^2\sigma^2$. The probability $P_B(\boldsymbol{x}_{ij})$, of choosing alternative $\boldsymbol{x}_{ij}$ from the set of alternatives $B = \{\boldsymbol{x}_{i1}, ...\boldsymbol{x}_{iJ}\}$, for the utility defined in equation (3.9) is

$$\begin{aligned} P_B(\boldsymbol{x}_{ij}) &= P[\lambda g(\boldsymbol{x}_{ik}) + \lambda\varepsilon_{ik} < \lambda g(\boldsymbol{x}_{ij}) + \lambda\varepsilon_{ij} \quad \text{for all} \quad k \neq j] \\ &= P[g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} \quad \text{for all} \quad k \neq j], \end{aligned} \tag{3.10}$$

which is the same as equation (3.5). The models in equations (3.2) and (3.9) are therefore equivalent. It is not possible to identify a value of λ; the likelihood is the same for any value of λ. For example, suppose $g(\boldsymbol{x}_{i1}) = 1$, $g(\boldsymbol{x}_{i2}) = 2$ and $\sigma^2 = 1$. The likelihood would be the same as the model where $g(\boldsymbol{x}_{i1} = 10$, $g(\boldsymbol{x}_{i2}) = 20$ and $\sigma^2 = 100$. Therefore, we cannot estimate unique joint estimates of $(g(.), \sigma^2)$ To be able to estimate unique utilities either the value of σ^2 or the

scale of $g(.)$ must be fixed. The standard way to normalise the scale of utility is to set the variance of the error term to 1. If $\lambda = \frac{1}{\sigma}$, the utility is written as

$$\frac{U_{ij}}{\sigma} = \frac{g(\boldsymbol{x}_{ij})}{\sigma} + \frac{\varepsilon_{ij}}{\sigma}, \quad j = 1, ..., J \tag{3.11}$$

where $Var\left(\frac{\varepsilon_{ij}}{\sigma}\right) = \frac{1}{\sigma^2} Var(\varepsilon_{ij}) = \frac{\sigma^2}{\sigma^2} = 1$. As the variance of the error terms is 1, there is only one possible scale of utilities.

Heteroscedastic errors

Normalisation of the utility scale is more complicated if the errors are not independent and identically distributed. Train (2003) discusses heteroscedastic errors, which involves the variance of the error terms being different for different segments of the population. Suppose the errors, ε_{ij}, are assumed to have different variances for male and female individuals. The model for utility is written as

$$\begin{aligned} U_{ij} &= g(\boldsymbol{x}_{ij}) + \varepsilon^f_{ij}, \quad \forall \text{ female } i \\ U_{ij} &= g(\boldsymbol{x}_{ij}) + \varepsilon^m_{ij}, \quad \forall \text{ male } i, \end{aligned} \tag{3.12}$$

where $Var(\varepsilon^f_{ij}) = \sigma^2_f$, $Var(\varepsilon^m_{ij}) = \sigma^2_m$ and $Var(\varepsilon^f_{ij}) \neq Var(\varepsilon^m_{ij})$. If we were to normalise as before, with the variances of the transformed errors set to 1 in each case, then the utilities for males and females would be on different scales. To fit a model including both male and female data, the overall utility scale must be the same. This can be set by normalising the variance of the error for females, and estimating the variance of the male errors relative to female errors.

Let $K = \frac{\sigma^2_m}{\sigma^2_f}$, which represents the variance of male utilities relative to female

utilities. Equation (3.12) can be written as

$$\begin{aligned} U_{ij} &= g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}^{f}, \quad \forall \quad \text{female} \quad i \\ \frac{U_{ij}}{\sqrt{K}} &= \frac{g(\boldsymbol{x}_{ij})}{\sqrt{K}} + \frac{\varepsilon_{ij}^{m}}{\sqrt{K}}, \quad \forall \quad \text{male} \quad i, \end{aligned} \tag{3.13}$$

where

$$\begin{aligned} Var\left(\frac{\varepsilon_{ij}^{m}}{\sqrt{K}}\right) &= \frac{1}{K} Var(\varepsilon_{ij}^{m}) \\ &= \frac{\sigma_f^2}{\sigma_m^2} \times \sigma_m^2 \\ &= \sigma_f^2 \\ &= Var(\varepsilon_{ij}^{f}). \end{aligned}$$

In equation (3.13) the utilities for male and females are now measured on the same scale by changing the variance of the males to be the same as females. Dividing the utilities of males $\sqrt{K}$ does not affect their choices. To complete the model, the variance of the female errors must be normalised to 1. The utility is

$$\begin{aligned} \frac{U_{ij}}{\sigma_f} &= \frac{g(\boldsymbol{x}_{ij})}{\sigma_f} + \varphi_{ij}^{f} \quad \forall \quad \text{female} \quad i \\ \frac{U_{ij}}{\sqrt{K}\sigma_f} &= \frac{g(\boldsymbol{x}_{ij})}{\sqrt{K}\sigma_f} + \varphi_{ij}^{m} \quad \forall \quad \text{male} \quad i, \end{aligned} \tag{3.14}$$

where $\varphi_{ij}^{f} = \dfrac{\varepsilon_{ij}^{f}}{\sigma_f}$, $\varphi_{ij}^{m} = \dfrac{\varepsilon_{ij}^{m}}{\sqrt{(K)}\sigma_f}$ and $Var(\varphi_{ij}^{f}) = Var(\varphi_{ij}^{m}) = 1$.

Correlated Errors

If errors are correlated over alternatives, normalising the variance of one alternative does not normalise the utility scale. The scale of utility differences must

be considered instead. Train (2003) considers an example where there are four alternatives. The utility of the four alternatives is given by

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \quad j = 1, ...4. \tag{3.15}$$

The error vector $\varepsilon_i = \varepsilon\{\varepsilon_{i1}, .., \varepsilon_{i4}\}$ has mean zero and covariance matrix

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ . & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ . & . & \sigma_{33} & \sigma_{34} \\ . & . & . & \sigma_{44} \end{pmatrix}. \tag{3.16}$$

Equation (3.15) can be written in terms of utility differences,

$$\tilde{U}_{ij1} = \tilde{g}(\boldsymbol{x}_{ij1}) + \tilde{\varepsilon}_{ij1}, \quad j = 2, 3, 4, \tag{3.17}$$

where $\tilde{U}_{ij1} = U_{ij} - U_{i1}$, $\tilde{g}(\boldsymbol{x}_{ij1}) = g(\boldsymbol{x}_{ij}) - g(\boldsymbol{x}_{i1})$ and $\tilde{\varepsilon_{ij1}} = \varepsilon_{ij} - \varepsilon_{i1}$. The vector of error differences $\tilde{\varepsilon}_{ij1} = \{\tilde{\varepsilon}_{i21}, \tilde{\varepsilon}_{i31}, \tilde{\varepsilon}_{i41}\}$ has mean zero and covariance matrix

$$\tilde{\Omega} = \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14} \\ . & \sigma_{11} + \sigma_{33} - 2\sigma_{13} & \sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14} \\ . & . & \sigma_{11} + \sigma_{44} - 2\sigma_{14} \end{pmatrix}. \tag{3.18}$$

A method to set the utility scale is to normalise the variance of one of the error difference. Setting the variance of an error difference sets the scale of utility differences and therefore sets the utility scale. To set the utility scale, suppose the variance of one error difference is normalised to, $Var(\tilde{\varepsilon_{i21}}) = 1$. Let $m = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$. After normalising the covariance matrix is

$$\Omega^* = \begin{pmatrix} 1 & \frac{\sigma_{11}+\sigma_{23}-\sigma_{12}-\sigma_{13}}{m} & \frac{\sigma_{11}+\sigma_{24}-\sigma_{12}-\sigma_{14}}{m} \\ . & \frac{\sigma_{11}+\sigma_{33}-2\sigma_{13}}{m} & \frac{\sigma_{11}+\sigma_{34}-\sigma_{13}-\sigma_{14}}{m} \\ . & . & \frac{\sigma_{11}+\sigma_{44}-2\sigma_{14}}{m} \end{pmatrix}.$$

The utility is now defined as

$$\frac{U_{ij}}{\sqrt{m}} = \frac{g(\boldsymbol{x}_{ij})}{\sqrt{m}} + \varepsilon^*, \tag{3.19}$$

where the new error, ε^*_{ij} has covariance Ω^*.

3.4.3 Property of Independence from irrelevant alternatives

Consider the set of alternatives $B = \{\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iJ}\}$. Let $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ be two alternatives in the set B. Define $P_B(\boldsymbol{x}_{ia})$ and $P_B(\boldsymbol{x}_{ib})$ to be the probability of choosing alternatives $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ from the set B respectively. The property of independence from irrelevant alternatives concerns the ratio of the two probabilities, $P_B(\boldsymbol{x}_{ia})$ and $P_B(\boldsymbol{x}_{ib})$:

$$\frac{P_B(\boldsymbol{x}_{ia})}{P_B(\boldsymbol{x}_{ib})}. \tag{3.20}$$

The property is discussed by Train (2003) and Luce (1959) and states that the ratio of the two probabilities, $P_B(\boldsymbol{x}_{ia})$ and $P_B(\boldsymbol{x}_{ib})$ depends on the two alternatives $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ and is independent of all other alternatives in set B. Some discrete choice models, such as the logit model satisfy this property whilst models such as the probit and mixed logit do not.

3.5 Logit Model

We now consider choices for the error distribution, starting with the type 1 extreme value distribution, which gives rise to the logit model. If a random variable X has a type 1 extreme value distribution the p.d.f. is given by

$$f_X(x) = \frac{1}{\sigma}\exp\left(\frac{-x+\mu}{\sigma}\right)\exp\left[-\exp\left(\frac{-x+\mu}{\sigma}\right)\right], \quad -\infty < x < \infty \tag{3.21}$$

where μ is the location parameter and σ is the scale parameter. The pdf $f_X(x)$ has mean $E(X) = \mu + 0.5722\sigma$ and variance $Var(X) = \frac{1}{6}\pi^2\sigma^2$. We require the errors to have a mean 0 (so that $g(\boldsymbol{x}_{ij})$ is interpreted as the population mean utility) and therefore $\mu = -0.5722\sigma$.

The cdf is given by

$$P(X < x) = \exp\left(-\exp\left(\frac{-x+\mu}{\sigma}\right)\right). \tag{3.22}$$

Figure (3.1) presents the probability distribution functions for the normal distribution and the type 1 extreme value distribution, where both distributions have a mean 0 and variance 1. The type 1 extreme value distribution has a positive skew. The type 1 extreme value distribution is often assumed for the errors because it presents a convenient form for the choice probability, i.e the logit model.

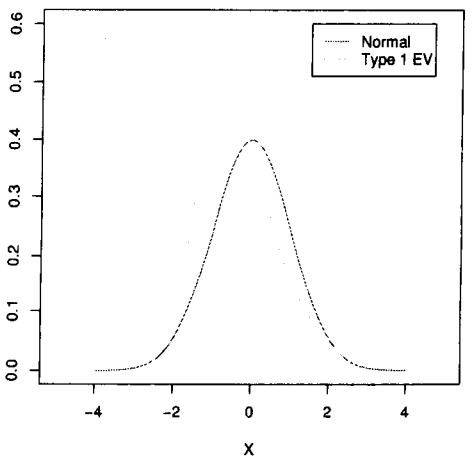


Figure 3.1: Probability distribution functions of Normal and type 1 Extreme value distributions

Suppose an individual i considers a set of possible alternatives $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iJ}\}$. The probability of choosing alternative $\boldsymbol{x}_{ij}$ in equation (3.3) can be written as

$$P_B(\boldsymbol{x}_{ij}) = P[\varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} - g(\boldsymbol{x}_{ik}) \quad \text{for all} \quad k \neq j] \tag{3.23}$$

The probability can be calculated using

$$\begin{aligned} P_B(\boldsymbol{x}_{ij}) &= \int P_B(\boldsymbol{x}_{ij}|\varepsilon_{ij}) f(\varepsilon_{ij}) d\varepsilon_{ij} \\ &= \int \prod_{k \neq j}^{J} P(\varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} - g(\boldsymbol{x}_{ik})|\varepsilon_{ij}) f(\varepsilon_{ij}) d\varepsilon_{ij}, \end{aligned} \tag{3.24}$$

The derivation of this probability for the logit model is shown by McFadden (1974). If ε_{ik} and ε_{ij} are assumed to have independent and identical type 1 extreme value distributions with scale parameter σ and location parameter $\mu =$

-0.5722σ then

$$f(\varepsilon_{ij}) = \frac{1}{\sigma}\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\exp\left[-\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\right] \tag{3.25}$$

and

$$P(\varepsilon_{ik} < g(\boldsymbol{x}_{ij})+\varepsilon_{ij}-g(\boldsymbol{x}_{ik})|\varepsilon_{ij}) = \exp\left(-\exp\left(\frac{-g(\boldsymbol{x}_{ij})-\varepsilon_{ij}+g(\boldsymbol{x}_{ik})-0.5722\sigma}{\sigma}\right)\right). \tag{3.26}$$

The probability $P_B(\boldsymbol{x}_{ij})$ is therefore

$$\begin{aligned}
P_B(\boldsymbol{x}_{ij}) &= \int_{\varepsilon_{ij}=-\infty}^{\varepsilon_{ij}=+\infty}\left\{\prod_{k\neq j}^{J}\exp\left[-\exp\left(\frac{-g(\boldsymbol{x}_{ij})-\varepsilon_{ij}+g(\boldsymbol{x}_{ik})-0.5722\sigma}{\sigma}\right)\right]\right. \\
&\quad \left.\times\frac{1}{\sigma}\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\exp\left[-\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\right]\right\}d\varepsilon_{ij} \\
&= \int_{\varepsilon_{ij}=-\infty}^{\varepsilon_{ij}=+\infty}\left\{\prod_{k=1}^{J}\exp\left[-\exp\left(\frac{-g(\boldsymbol{x}_{ij})-\varepsilon_{ij}+g(\boldsymbol{x}_{ik})-0.5722\sigma}{\sigma}\right)\right]\right. \\
&\quad \left.\times\frac{1}{\sigma}\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\right\}d\varepsilon_{ij} \\
&= \int_{\varepsilon_{ij}=-\infty}^{\varepsilon_{ij}=+\infty}\left\{\exp\left[-\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\sum_{k=1}^{J}\exp\left(\frac{g(\boldsymbol{x}_{ik})-g(\boldsymbol{x}_{ij})}{\sigma}\right)\right]\right. \\
&\quad \left.\times\frac{1}{\sigma}\exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)\right\}d\varepsilon_{ij}.
\end{aligned} \tag{3.27}$$

The probability, $P_B(\boldsymbol{x}_{ij})$, can be derived using substitution. If $a = \sum_{k=1}^{J}\exp\left(\frac{g(\boldsymbol{x}_{ik})-g(\boldsymbol{x}_{ij})}{\sigma}\right.$ and $z = \exp\left(\frac{-\varepsilon_{ij}-0.5722\sigma}{\sigma}\right)$ then

$$
\begin{aligned}
P_B(\boldsymbol{x}_{ij}) &= \int_{\infty}^{0} -\exp(-az)dz \\
&= \left[\frac{1}{a}\exp(-az)\right]_{\infty}^{0} \\
&= \frac{1}{a} \\
&= \left(\frac{\exp\left(\frac{g(\boldsymbol{x}_{ij})}{\sigma}\right)}{\sum_{k=1}^{J}\exp\left(\frac{g(\boldsymbol{x}_{ik})}{\sigma}\right)}\right)
\end{aligned}
\tag{3.28}
$$

Equation (3.28) is called the logit model.

3.5.1 Normalising the utility scale

Section 3.4.2 discusses generally how a utility scale can be normalised. This now needs to be applied to the logit model. The probability in equation (3.28) is for an undefined utility scale so different values of each parameter can give the same probability. The utilities are measured relative to the utilities of two fixed alternatives, the best possible alternative and another alternative considered worse than most alternatives. Suppose the utility of the lower alternative is given a value D and the utility of the best alternative is given a value C. The mean utility $g(\boldsymbol{x}_{ij})$ can be written as $g(\boldsymbol{x}_{ij}) = C - f(\boldsymbol{x}_{ij})$ where $f(\boldsymbol{x}_{ij})$ is the mean decrease in utility from the best alternative. If individual i chooses alternative $\boldsymbol{x}_{ij}$ from the set of alternative B, which includes the lower of the two alternatives,

then the following condition must be satisfied,

$$g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} > D$$
$$\varepsilon_{ij} > -g(\boldsymbol{x}_{ij}) + D. \tag{3.29}$$

The probability in equation (3.24) can now be written to include this condition as

$$P_B(\boldsymbol{x}_{ij}) = \int \prod_{k \neq j}^{J} P(\varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} - g(\boldsymbol{x}_{ik}) | \varepsilon_{ij}) I(\varepsilon_{ij} > -g(\boldsymbol{x}_{ij}) + D | \varepsilon_{ij}) f(\varepsilon_{ij}) d\varepsilon_{ij}, \tag{3.30}$$

where $I(\varepsilon_{ij} > -g(\boldsymbol{x}_{ij}) + D|\varepsilon_{ij}) = 1$ if $\varepsilon_{ij} > -g(\boldsymbol{x}_{ij}) + D$ and 0 otherwise. The probability is then calculated using

$$P_B(\boldsymbol{x}_{ij}) = \int_{-g(\boldsymbol{x}_{ij})+C}^{\infty} \prod_{k \neq j}^{J} P(\varepsilon_{ik} < g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} - g(\boldsymbol{x}_{ik}) | \varepsilon_{ij}) f(\varepsilon_{ij}) d\varepsilon_{ij}. \tag{3.31}$$

It can be shown that the probability, $P_B(\boldsymbol{x}_{ij})$, is now equal to

$$P_B(\boldsymbol{x}_{ij}) = \left(\frac{\exp\left(\frac{g(\boldsymbol{x}_{ij})}{\sigma}\right)}{\sum_{k=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ik})}{\sigma}\right)} \right) \times \left(1 - \exp\left(-\frac{\exp\left(\frac{g(\boldsymbol{x}_{ij})}{\sigma}\right)}{\sum_{k=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ik})}{\sigma}\right)} \exp\left(\frac{g(\boldsymbol{x}_{ij}) - D - 0.5722\sigma}{\sigma} \right) \right) \right). \tag{3.32}$$

3.5.2 RANDOM EFFECTS IN LOGIT MODELS

In a logit model, the errors are assumed to be independently and identically distributed. However, individuals' utilities can vary with unobserved variables or randomly. Consider two similar alternatives, $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ that are defined using 5 attributes. Suppose alternative $\boldsymbol{x}_{ia}$ is at level 3 for each of the five attributes, and alternative $\boldsymbol{x}_{ib}$ is at level 3 for four attributes and level 4 for one attribute. If the utility individual i has for alternative $\boldsymbol{x}_{ia}$ is less than the population mean utility for alternative $\boldsymbol{x}_{ia}$, then we would expect that the utility individual i has for alternative $\boldsymbol{x}_{ib}$ is also less than the population mean utility for alternative $\boldsymbol{x}_{ib}$. Consequently, we would judge the errors associated with the two alternatives, $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$, to be correlated, and the errors to not be independent and identically distributed.

To account for the correlation, a random effect can included in the model. A random effect can be additive or multiplicative. Train (2003) discusses the multiplicative random effect. Define α_i to be a multiplicative random effect. The utility is then defined as

$$U_{ij} = \alpha_i g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \quad j = 1, ..., J, \tag{3.33}$$

where the random effect, α_i, represents the mean deviation of individual i's utilities from the population mean utilities.

In a logit model the random effect can be dependent on a covariate, for example age. Let $\alpha_i = \rho y_i$, where y_i is the age of individual i. Equation (3.33) can now be written as

$$U_{ij} = \rho y_i g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}. \tag{3.34}$$

Other specifications can be considered for the random effect, such as $\alpha_i = \rho y_i + \varphi y_i^2$, where ρ is positive and φ is negative. As age increases, utility increases but at a decreasing rate.

Logit models can include parameters for each individual that vary over observed variables such as age. However, parameters that vary with respect to unobserved variables or randomly cannot be included in a logit model. Suppose the random effect is $\alpha_i = \rho y_i + \mu_i$, where μ_i is a random variable. Substituting into Equation (3.33) gives

$$\begin{aligned} U_{ij} &= (\rho y_i + \mu_i) g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} \\ &= \rho y_i g(\boldsymbol{x}_{ij}) + \mu_i g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \end{aligned} \tag{3.35}$$

which can be written as

$$U_{ij} = \rho y_i g(\boldsymbol{x}_{ij}) + \tilde{\varepsilon_{ij}}, \tag{3.36}$$

where $\tilde{\varepsilon_{ij}} = \mu_i g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}$. The new error, $\tilde{\varepsilon_{ij}}$, cannot be independently and identically distributed. The random variable, μ_i, will be in the utility of each alternative for individual i and therefore the errors are correlated over individual i. The variance of the error, $\tilde{\varepsilon_{ij}}$ is

$$Var(\tilde{\varepsilon_{ij}}) = Var(\mu_i) g(\boldsymbol{x}_{ij})^2 + Var(\varepsilon_{ij}).$$

As the variance of $\tilde{\varepsilon}_{ij}$ depends on $\boldsymbol{x}_{ij}$ and μ_i, there is a different variance for each alternative and individual. Therefore the errors are not identically distributed. As the errors, $\tilde{\varepsilon}_{ij}$ are not independent and have different variances, a logit model is not suitable for the utiltiy defined in equation (3.35).

3.5.3 THE LOGIT MODEL AND INDEPENDENCE FROM IRRELEVANT ALTERNATIVES

The logit model satisfies the property of independence from irrelevant alternatives. This is demonstrated in Train (2003) and can be described as follows. Let $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ be two alternatives in the set B. Define $P_B(\boldsymbol{x}_{ia})$ and $P_B(\boldsymbol{x}_{ib})$ as the probability of choosing alternative $\boldsymbol{x}_{ia}$ and alternative $\boldsymbol{x}_{ib}$ from the set B respectively. The ratio of the logit probabilities is

$$\begin{aligned}
\frac{P_B(\boldsymbol{x}_{ia})}{P_B(\boldsymbol{x}_{ib})} &= \frac{\left(\dfrac{\exp(g(\boldsymbol{x}_{ia}))}{\sum_{j=1}^{J}\exp(g(\boldsymbol{x}_{ij}))}\right)}{\left(\dfrac{\exp(g(\boldsymbol{x}_{ib}))}{\sum_{j=1}^{J}\exp(g(\boldsymbol{x}_{ij}))}\right)} \\
&= \frac{\exp(g(\boldsymbol{x}_{ia}))}{\exp(g(\boldsymbol{x}_{ib}))} \qquad (3.37) \\
&= \exp[g(\boldsymbol{x}_{ia}) - g(\boldsymbol{x}_{ib})].
\end{aligned}$$

The ratio of the two probabilities only depends on alternatives $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$. Therefore the relative odds of choosing alternative $\boldsymbol{x}_{ia}$ over alternative $\boldsymbol{x}_{ib}$ is independent of any other alternatives in set B.

3.6 Probit Model

The utility individual i has for alternative $\boldsymbol{x}_{ij}$ is described by Equation (3.2). Define $\boldsymbol{\varepsilon}_i = \{\varepsilon_{i1}, ..., \varepsilon_{iJ}\}$ to be the vector of errors for individual i corresponding to each alternative in the set $B = \{\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iJ}\}$. In the probit model the distribution of ε_i is assumed to be $\boldsymbol{\varepsilon}_i \sim N(0, \Omega_i)$, where Ω_i is the covariance matrix

for individual i. Define A_{ij} to be the set of error terms that result in individual i choosing alternative $\boldsymbol{x}_{ij}$:

$$A_{ij} = \{\varepsilon_i \quad s.t. \quad g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} > g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} \quad \forall j \neq k\}. \tag{3.38}$$

The probability, $P_B(\boldsymbol{x}_{ij})$, of choosing alternative $\boldsymbol{x}_{ij}$ from the set $B = \{\boldsymbol{x}_{i1}, ..\boldsymbol{x}_{iJ}\}$ is

$$P_B(\boldsymbol{x}_{ij}) = \int_{\varepsilon_i \in A_{ij}} \varphi(\varepsilon_i) d\varepsilon_i, \tag{3.39}$$

where

$$\varphi(\varepsilon_i) = \frac{1}{(2\pi)^{J/2}|\Omega_i|^{1/2}} \exp\left(\frac{\varepsilon_i' \Omega_i^{-1} \varepsilon_i}{2}\right).$$

It is easier to show this probability for a set with two alternatives. If an individual i has a set of possible alternatives $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$, then from equation (3.3) the probability of choosing one of the pair of alternatives, $P_B(\boldsymbol{x}_{i1})$, can be written as

$$P_B(\boldsymbol{x}_{i1}) = P[\varepsilon_{i2} - \varepsilon_{i1} < g(\boldsymbol{x}_{i1}) - g(\boldsymbol{x}_{i2})]. \tag{3.40}$$

If the errors ε_{i1} and ε_{i2} are assumed to be independent and identically distributed with normal distributions,

$$\varepsilon_{i1} \sim N(0, \sigma^2), \qquad \varepsilon_{i2} \sim N(0, \sigma^2), \tag{3.41}$$

then the difference $\varepsilon = \varepsilon_{i2} - \varepsilon_{i1}$ also has a normal distribution, $\varepsilon \sim N(0, 2\sigma^2)$. Therefore, the probability $P_B(\boldsymbol{x}_{i1})$ is given by

$$P_B(\boldsymbol{x}_{i1}) = \Phi\left(\frac{g(\boldsymbol{x}_{i1}) - g(\boldsymbol{x}_{i2})}{\sqrt{2\sigma^2}}\right), \tag{3.42}$$

where Φ is the standard normal cdf. Equation (3.42) is called the probit model.

To normalise the utility scale suppose $\boldsymbol{x}_{i2}$ is the lower alternative with fixed utility D then the probability is written as

$$P_B(\boldsymbol{x}_{i1}) = P[\varepsilon_{i1} > -g(\boldsymbol{x}_{i1}) + D]. \tag{3.43}$$

If ε_{i1} is assumed to have a normal distribution, $\varepsilon_{i1} \sim N(0, \sigma^2)$, the probability, $P_B(\boldsymbol{x}_{i1})$ is given by

$$P_B(\boldsymbol{x}_{i1}) = \Phi\left(\frac{g(\boldsymbol{x}_{i1}) - D}{\sigma}\right), \tag{3.44}$$

where Φ is the standard normal cdf.

3.6.1 Properties of Probit Model

Probit models allow individuals' utilities to be correlated. A limitation of the probit model is that a Normal distribution is required for all unobserved components of utility. The method of normalising the utility scale for correlated errors is described in Train (2003) and discussed in Section 3.4.

Probit models can include random coefficients which are normally distributed. Train (2003) considers an example where utility is assumed to be linear in parameters and the parameters vary randomly over individuals. The utility, U_{ij}, is defined as

$$U_{ij} = \boldsymbol{\theta}_i^T \boldsymbol{x}_{ij} + \varepsilon_{ij}, \quad j = 1, ..., J, \tag{3.45}$$

where $\boldsymbol{\theta}_i$ is the vector of parameters for individual i. Suppose the parameter vector, $\boldsymbol{\theta}_i$, is normally distributed with mean $\boldsymbol{b}$ and covariance $\boldsymbol{W}$: $\boldsymbol{\theta}_i \sim N(\boldsymbol{b}, \boldsymbol{W})$. The utility in equation(3.45) can be written as

$$U_{ij} = \boldsymbol{b}^T \boldsymbol{x}_{ij} + \mu_{ij}, \tag{3.46}$$

where $\mu_{ij} = \tilde{\boldsymbol{\theta}}_i^T \boldsymbol{x}_{ij} + \varepsilon_{ij}$ and $\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i - \boldsymbol{b}$ is the deviation of individual i from the population mean utility.

Train (2003) describes the covariance of μ_{ij} using an example with two alternatives. The utilities of alternatives $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ are

$$\begin{aligned} U_{i1} &= \boldsymbol{\theta}_i \boldsymbol{x}_{i1} + \varepsilon_{i1} \\ U_{i2} &= \boldsymbol{\theta}_i \boldsymbol{x}_{i2} + \varepsilon_{i2}, \end{aligned} \tag{3.47}$$

where $\boldsymbol{\theta}_i \sim N(\boldsymbol{b}, \sigma_B)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$, $\quad j = 1, 2$. The utilities can be written as

$$\begin{aligned} U_{i1} &= \boldsymbol{b}\boldsymbol{x}_{i1} + \mu_{i1} \\ U_{i2} &= \boldsymbol{b}\boldsymbol{x}_{i2} + \mu_{i2}, . \end{aligned} \tag{3.48}$$

The distribution of the vector $\boldsymbol{\mu}_i = \{\mu_{i1}, \mu_{i2}\}$ is $\boldsymbol{\mu}_i \sim N(0, \Omega)$, where

$$\Omega = \sigma_B \begin{pmatrix} \boldsymbol{x}_{i1}^2 & \boldsymbol{x}_{i1}\boldsymbol{x}_{i2} \\ \boldsymbol{x}_{i1}\boldsymbol{x}_{i2} & \boldsymbol{x}_{i2}^2 \end{pmatrix} + \sigma_\varepsilon \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.49}$$

To set the scale of utility, let $\sigma_\varepsilon = 1$, so Ω becomes

$$\Omega = \sigma_B \begin{pmatrix} \boldsymbol{x}_{i1}^2 & \boldsymbol{x}_{i1}\boldsymbol{x}_{i2} \\ \boldsymbol{x}_{i1}\boldsymbol{x}_{i2} & \boldsymbol{x}_{i2}^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.50}$$

3.6.2 The Probit Model and independence from irrelevant alternatives

A probit model does not satisfy the property of independence from irrelevant alternatives. This ratio of the probabilities of choosing alternatives $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$ is given by

$$\frac{P_B(\boldsymbol{x}_{ia})}{P_B(\boldsymbol{x}_{ib})} = \frac{\int\limits_{\varepsilon_i \in A_{ia}} \varphi(\varepsilon_i) d\varepsilon_i}{\int\limits_{\varepsilon_i \in A_{ib}} \varphi(\varepsilon_i) d\varepsilon_i}, \tag{3.51}$$

where $A_{ia} = \{\varepsilon_i \quad s.t. \quad g(\boldsymbol{x}_{ia}) + \varepsilon_{ia} > g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} \quad \forall j \neq k\}$, $A_{ib} = \{\varepsilon_i \quad s.t. \quad g(\boldsymbol{x}_{ib}) + \varepsilon_{ib} > g(\boldsymbol{x}_{ik}) + \varepsilon_{ik} \quad \forall j \neq k\}$ and

$$\varphi(\varepsilon_i) = \frac{1}{(2\pi)^{J/2} |\Omega_i|^{1/2}} \exp\left(\frac{\varepsilon_i' \Omega_i^{-1} \varepsilon_i}{2}\right).$$

Equation (3.51) cannot be simplified. Therefore the ratio of probabilities of the two alternatives depends on all the other alternatives and the probit model does not satisfy the property of independence from irrelevant alternatives.

3.7 Mixed logit model

Suppose the mean utility, $g(\boldsymbol{x}_{ij})$, in equation (3.2) is defined as $g(\boldsymbol{x}_{ij}) = \boldsymbol{\theta}^T \boldsymbol{x}_{ij}$, where the alternative $\boldsymbol{x}_{ij}$ is defined using dummy variables and $\boldsymbol{\theta}$ is the vector of corresponding parameters. The utility is then defined as

$$U_{ij} = \boldsymbol{\theta}\, \boldsymbol{x}_{ij} + \varepsilon_{ij}. \tag{3.52}$$

In a mixed logit model the errors, ε_{ij}, are assumed to have an independent and identically type 1 extreme value distribution and parameters $\boldsymbol{\theta}$ are allowed to vary over individual. Define $\boldsymbol{\theta}_i$ to be the set of parameters for individual i, where $\boldsymbol{\theta}_i$ is allowed to have any distribution with mean $\boldsymbol{b}$ and covariance $\boldsymbol{W}$.

The utility using a mixed logit model is written as

$$U_{ij} = \boldsymbol{\theta}_i \boldsymbol{x}_{ij} + \varepsilon_{ij}. \tag{3.53}$$

Define $P_B(\boldsymbol{x}_{ij}|\boldsymbol{\theta}_i)$ to be the probability of choosing alternative $\boldsymbol{x}_{ij}$ from the set of alternatives $B = \{\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iJ}\}$, evaluated at the set of parameters $\boldsymbol{\theta}_i$. Define $f(\boldsymbol{\theta}_i)$ to be the pdf of $\boldsymbol{\theta}$. The probability, $P_B(\boldsymbol{x}_{ij})$, is calculated using the mixed logit model by

$$P_B(\boldsymbol{x}_{ij}) = \int_{i=1}^{I} P_B(\boldsymbol{x}_{ij}|\boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \tag{3.54}$$

where

$$P_B(\boldsymbol{x}_{ij}|\boldsymbol{\theta}_i) = \frac{\exp(\boldsymbol{\theta}_i^T \boldsymbol{x}_{ij})}{\sum_{k=1}^{J} \exp(\boldsymbol{\theta}_i \boldsymbol{x}_{ik})}.$$

The probability in equation (3.54) is a weighted average of the logit function evaluated at different values of $\boldsymbol{\theta}_i$, where the weights are given by the pdf $f(\boldsymbol{\theta}_i)$. For a set of pairwise choices $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$ the probability $P_B(\boldsymbol{x}_{i1})$ is calculated using the mixed logit model by

$$P_B(\boldsymbol{x}_{i1}) = \int P_B(\boldsymbol{x}_{i1}|\boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \tag{3.55}$$

where

$$P_B(\boldsymbol{x}_{i1}|\boldsymbol{\theta}_i) = \frac{\exp(\boldsymbol{\theta}_i^T \boldsymbol{x}_{i1})}{\exp(\boldsymbol{\theta}_i^T \boldsymbol{x}_{i1}) + \exp(\boldsymbol{\theta}_i^T \boldsymbol{x}_{i2})}.$$

3.7.1 The mixed logit model and independence from irrelevant alternatives

A mixed logit model does not satisfy the property of independence from irrelevant alternatives. This can be shown by considering the ratio of the probabilities of

choosing alternatives $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$:

$$\frac{P_B(\boldsymbol{x}_{ia})}{P_B(\boldsymbol{x}_{ib})} = \frac{\int \frac{\exp(g(\boldsymbol{x}_{ia}))}{\sum_{j=1}^{J} \exp(g(\boldsymbol{x}_{ij}))} f(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \frac{\exp(g(\boldsymbol{x}_{ib}))}{\sum_{j=1}^{J} \exp(g(\boldsymbol{x}_{ij}))} f(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{3.56}$$

The denominators of the two probabilities in Equation (3.56) do not cancel. The ratio of probabilities of the two alternatives depends on all the other alternatives. Therefore the relative odds of choosing alternative $\boldsymbol{x}_{ia}$ over alternative $\boldsymbol{x}_{ib}$ is not independent of the other alternatives in the set B.

3.8 Discrete choice experiments in health care

This section reviews examples of where discrete choice experiments have been used to estimate utilities in health care. Ryan et al. (2001) and Ryan (2004) are examples of articles which review the use of discrete choice data in health care. Discrete choice experiments were introduced into Health Economics to consider other factors of health care in addition to the QALY. These factors include waiting time, location of treatment, type of care and type of staff providing care. Discrete choice experiments allow trade-offs between such differences in health care. Discrete choice experiments are based on the assumption that health care interventions or services can be described by their characteristics (or attributes) and that an individual's valuation depends on the levels of these characteristics.

Ryan and Farrar (2000) describes the process of conducting a discrete choice experiment, which consists of five stages. Initially the characteristics describing the health intervention or service are identified. A number of possible levels are then defined for each characteristic. Each possible health care scenario can then be

listed by combining each possible characteristic level. A discrete choice experiment involves collecting preferences on a subset of these scenarios. A respondent is presented with a set of possible scenarios and asked to choice the preferred option. This is carried out for several groups. This type of data is called discrete choice data. A model is then fitted to the data in order to estimate a value of every possible scenario. Models used are often the logit or probit model. The interpretation of the resulting estimate of each scenario is determined by the purpose of the discrete choice experiment. This thesis is considering using discrete choice data to estimate utilities for use in deriving QALYs. Therefore model examples presented in this section can be used for this purpose.

Salomon (2003), using the EQ-5D data in Dolan (1997) discusses the analysis of ordinal rank data. The model proposed is

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij} \tag{3.57}$$

where $g(\boldsymbol{x}_{ij})$ is the population mean utility for health state $\boldsymbol{x}_{ij}$ and is given by the function $g(x_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\theta}$ The vector $\boldsymbol{x}_{ij}$ consists of dummy variables representing the levels of each of the five attributes in the EQ-5D classification system defining the health state, and $\boldsymbol{\theta}$ is the vector of unknown parameters. A logit model was used to analyse the data.

The model in Equation (3.57) does not produce utilities on the required scale where full health has a utility of 1 and death has a utility of 0. Possible methods for scaling the utilities were discussed. The model in Equation (3.57) can be written as

$$U_{ij} = \lambda(g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}) + \delta \tag{3.58}$$

where λ is the normalising constant and δ is the value of the best possible health state. The likelihood function will be the same for all values of α and δ where $\alpha, \delta \neq 0$. The best health state is always full health and as δ corresponds to the utility of the best possible health state the value of δ is always $\delta = 1$.

Three alternatives for λ were considered. The first alternative proposed was to transform the scale such that the worst possible health state defined by the EQ-5D classification system has a utility equal to the mean observed TTO value for that state, denoted $\overline{TTO}_{33333}$. The modelled value of the worst state equals the sum of all the parameter estimates, $\sum \hat{\theta}$. Therefore

$$\lambda \sum \hat{\theta} + 1 = \overline{TTO}_{33333}, \tag{3.59}$$

and λ is:

$$\lambda = \frac{\overline{TTO}_{33333} - 1}{\sum \hat{\theta}}. \tag{3.60}$$

The second alternative for λ defines a scale where the utility of the worst state is 0. In this case

$$\lambda \sum \hat{\theta} + 1 = 0, \tag{3.61}$$

and λ is

$$\hat{\lambda} = \frac{-1}{\sum \hat{\theta}}. \tag{3.62}$$

In the final alternative for λ, the model in Equation (3.58) is written as

$$U_{ij} = \lambda(g(\boldsymbol{x}_{ij}) + D + \varepsilon_{ij}) + \delta, \tag{3.63}$$

where D is a dummy variable representing the health state death, and is equal to 1 if the health state is death and is equal to 0 otherwise. If the utility of death is given a value of 0 for all individuals then,

$$\lambda D + 1 = 0, \tag{3.64}$$

and λ is

$$\lambda = \frac{-1}{D}. \tag{3.65}$$

A similar model is used by McCabe et al. (2006) where a rank model is fitted to rank data collected using both the HUI2 and SF-6D classification systems. The estimated utilities are then re-scaled using equation (3.65) to be defined on a scale where death has a utility of 0 and perfect health has a utility of 1. The utilities estimated using rank data are compared with utilities derived using the Standard Gamble method.

Lancsar et al. (2003) discuss a study that involves four discrete choice experiments (DCEs) undertaken within a multi-centre, randomised cross-over, controlled trial of three preventive asthma medications. The trial had a two period double blind crossover design for two drugs, Montelukast and Formoterol, and a follow-up period on a third, Fluticasone. Patients completed four DCE questionnaires. The first was completed on entry to the trial before the patients received any trial medication. The other three were completed after each of the three treatment phases. A classification system with 10 attributes was used. Each attribute had between 2 and 4 levels, defining 131,072 possible scenarios. There were 16 versions of the DCE questionnaire, each including 28 scenarios. A total of 58 patients were randomised to complete one questionnaire at each of the four times specified.

For each of the 28 scenarios in the survey assigned to them, patients choose

which treatment option they would prefer from the following set of options:

1 Medication they had been taking for the previous 6 weeks,

2 Medication described in the scenario, or

3 No preventive medication.

The model specified in Lancsar et al. (2003) is for the DCE questionnaire completed at the first visit. It is assumed that individual i has a utility, U_{ij} for state $\boldsymbol{x}_{ij}$ shown by

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \tag{3.66}$$

where

$$g(\boldsymbol{x}_{ij}) = \theta_j^T X_{ij} + \gamma_i^T Z_i.$$

X is a vector of attributes describing the asthma medications containing either values describing the current medication (for j=1), the design variables for the scenario being valued (for j=2) or zeros when no medication is preferred (for j=3). The vector Z contains characteristic variables on individual subjects. A logit model is fitted to the data. If the data from all four DCEs were included in the analysis, Lancsar et al. (2003) concluded it may not be reasonable to assume that the ε_{ij} are independent and instead may have a complex correlation structure.

? analyses the same data as Lancsar et al. (2003) using a mixed logit model, which allows the parameters to vary over individuals. The utility individual i has for alternative j in scenario s is given by

$$U_{isj} = \boldsymbol{x}_{isj}^T \boldsymbol{\theta}_i + \varepsilon_{isj}, \tag{3.67}$$

where $\boldsymbol{x}_{isj}$ is a vector of attributes describing the asthma medications for scenario s and containing either values describing the current medication (for j=1), the design variables for the scenario being valued (for j=2) or zeros when no medication is specified (for j=3)

Farrar et al. (2000) investigates the use of discrete choice models (DCM) to elicit the views of planners of health care, providers of health care and consumers in the area of priority setting. Priority setting involves choosing between competing demands on the health care budget. The hospital trust used had an aim to gain as much benefit as possible from the choice of clinical service developments and wanted the measure of benefit used to reflect the preferences of consultants working at the hospital

A classification system was developed which had five dimensions: the level of evidence of clinical effectiveness; the size of health gain; the developments contribution to professional development; the developments contribution to education, training and research; and the strategy area, which is whether the proposed development represents a local and/or national priority. Each dimension was assigned a number of levels. The total number of possible scenarios defined by the dimensions and levels is 216. Sixteen scenarios were identified which could be used in a DCE questionnaire and allow benefit or utility scores to be estimated for all possible scenarios or clinical service developments. The 16 scenarios were paired into 8 choices determined on the basis that each choice included clear trade-offs between dimensions of benefit. The 216 consultants working within the hospital trust were asked to complete a questionnaire. As multiple observations were obtained from each individual the random effects probit model was used to analyse the data.

The model used is similar to the models described previously in this section

except that utility differences are used. The model is written in the form

$$V_{ijk} = h(\boldsymbol{x}_{ijk}) + \varepsilon + \alpha, \tag{3.68}$$

where V_{ijk} is the change in utility from development $\boldsymbol{x}_{ij}$ to development $\boldsymbol{x}_{ik}$, and $h(\boldsymbol{x}_{ijk}) = \boldsymbol{\theta}^T(\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik})$. It is unclear what ε and α represent. They are defined as the unobservable error terms where $corr[\varepsilon, \alpha] = \rho$ and ρ takes account of any correlation between observations from any one individual. It might be assumed that ε was the error term and α was the random effect. However as the model represents the difference in utilities the random effect would cancel out, and the error term is independent and cannot be correlated with another error. The two terms could represent the utility of the two utilities before the difference is taken.

Burr et al. (2007) use a discrete choice experiment to estimate the utilities for health outcomes resulting from glaucoma. A classification system with 6 dimensions was developed. These dimensions were: central and near vision; lighting and glare; mobility; activities of daily living; eye discomfort and other effects. Each dimension had four levels of severity, ranging from no difficulty, given a value of 0, to severe difficulty, given a value of 3. The classification system defined 4096 health states. Thirty two health states were selected and then the mirror image of each found by changing the levels in each health state, i.e. $0 \Rightarrow 1$, $1 \Rightarrow 2$, $2 \Rightarrow 3$ and $3 \Rightarrow 0$. For example health state 0123 becomes health state 1230. Each of the two health states are then used as a pairwise comparison. There are then thirty two pairwise comparisons used in a questionnaire sent out to glaucoma patients. A conditional logistic regression model is used to analyse the response data. A Wald test was used to test for evidence of a significant difference between levels for each dimension. If there is no difference, the levels were combined and the model re-estimated.

3.9 Bayesian Inference for discrete choice data

Albert and Chib (1993) develop Bayesian methods for modeling categorical response data using data augmentation. The method is described using the probit model as an example. The probit regression model for binary outcomes has an underlying normal structure on the latent variables. In data augmentation, values of the latent variables are simulated from truncated normal distributions. If the latent data are known, the posterior distribution of the parameters can be derived. New values of latent data can then be sampled from the posterior distributions.

Suppose that $Y_1, ...Y_N$ are observed independent binary random variables where $Y_i, \quad i = 1, ...N$ has a Bernoulli distribution with probability of success p_i. Define the binary regression model as

$$p_i = H(\boldsymbol{x}_i^T \boldsymbol{\theta}), \tag{3.69}$$

where $\boldsymbol{\theta}$ is a $k \times 1$ vector of unknown parameters, and $\boldsymbol{x}_i^T = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iN})$ is a vector of known continuous or discrete covariates. The function H is a known cdf linking the probabilities p_i with the linear structure $\boldsymbol{x}_i^T \boldsymbol{\theta}$. Define $\pi(\boldsymbol{\theta})$ to be the prior distribution of parameter $\boldsymbol{\theta}$. The posterior distribution is given by

$$\frac{\pi(\boldsymbol{\theta}) \prod_{i=1}^{N} H(\boldsymbol{x}_i^T \boldsymbol{\theta})^{y_i} (1 - H(\boldsymbol{x}_i^T \boldsymbol{\theta}))^{1-y_i}}{\int \pi(\boldsymbol{\theta}) \prod_{i=1}^{N} H(\boldsymbol{x}_i^T \boldsymbol{\theta})^{y_i} (1 - H(\boldsymbol{x}_i^T \boldsymbol{\theta}))^{1-y_i} d\boldsymbol{\theta}}. \tag{3.70}$$

To describe the method of data augmentation, Albert and Chib (1993) let $H = \Phi$, leading to a probit model. Suppose $Z_1, ..., Z_N$ are latent variables where each

Z_i are independent $N(\boldsymbol{x}_i^T\boldsymbol{\theta}, 1)$. Define Y_i to be a dummy variable where

$$Y_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.71}$$

Each Y_i is an independent Bernoulli random variables with $p_i = P(Y_i = 1) = \Phi(\boldsymbol{x}_i^T\boldsymbol{\theta})$. The joint posterior of $\boldsymbol{\theta}$ and $\boldsymbol{Z} = (Z_1, ..., Z_N)$ given the data $\boldsymbol{Y} = (Y_1, ..., Y_N)$ is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{Z}|\boldsymbol{y}) &= C\pi(\boldsymbol{\theta}) \left\{ \prod_{i=1}^{N} \{I(Z_i > 0)I(y_i = 1) + (Z_i \leq 0)I(y_i = 0)\} \right. \\ &\quad \left. \times \Phi(Z_i; \boldsymbol{x}_i^T\boldsymbol{\theta}, 1) \right\}, \end{aligned} \tag{3.72}$$

where $\Phi(; \boldsymbol{x}_i^T\boldsymbol{\theta}, 1)$ is the $N(\boldsymbol{x}_i^T\boldsymbol{\theta}, 1)$ cdf, $I(X \epsilon A)$ is the indicator function that is equal to 1 if the random variable X is contained in the set A, and C is a proportionality constant. The posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{Z}$ is given by

$$\pi(\boldsymbol{\theta}|y, \boldsymbol{Z}) = C\pi(\boldsymbol{\theta}) \prod_{i=1}^{N} \Phi(Z_i; \boldsymbol{x}_i^T\boldsymbol{\theta}, 1). \tag{3.73}$$

The posterior distribution of each $\boldsymbol{Z}_i$ given $\boldsymbol{\theta}$ is independent with

$$Z_i|\boldsymbol{y}, \boldsymbol{\theta} \sim N(\boldsymbol{x}_i^T\boldsymbol{\theta}, 1) \begin{cases} \text{truncated at the left by} & 0 \text{ if } y_i = 1 \\ \text{truncated at the right by} & 0 \text{ if } y_i = 0 \end{cases} \tag{3.74}$$

Albert and Chib (1993) then discuss how it is possible to generalise the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{Z})$, by applying suitable mixtures of normal distributions. This generalisation allows investigation into the sensitivity of the fitted probabilities to the choice of link function. The probit link can be generalised by choosing the link pdf to be the family of t distributions.

Let the Z_i be independently distributed from t distributions with location $\boldsymbol{x}_i^T\boldsymbol{\theta}$, scale parameter 1 and degrees of freedom ν. Introducing, an additional variable, λ_i, the distribution of Z_i is now written as the following scale mixture of a normal distribution: $Z_i \sim N(\boldsymbol{x}_i^T\boldsymbol{\theta}, \lambda_i^{-1})$ and $\lambda_i \sim Gamma(\nu/2, 2/\nu)$, where the gamma pdf is proportional to $\lambda_i^{\nu/2-1}\exp(-\nu\lambda_i/2)$.

Let $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_N)$ be the vector of scale parameters and suppose the parameter $\boldsymbol{\theta}$ has a uniform prior distribution. The posterior for $\boldsymbol{Z}$, $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ and ν is

$$\begin{aligned}\pi(\boldsymbol{Z}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \nu|\boldsymbol{y}) \;=\; & C\pi(\nu) \times \\ & \prod_{i=1}^{N}\left\{\sqrt{\frac{\lambda_i}{2\pi}} \times \{I(Z_i > 0)I(Y_i = 1) + I(Z_i \leq 0)I(Y_i = 0)\}\right. \\ & \left.\times \exp\left(\frac{-\lambda}{2}(Z_i - \boldsymbol{x}_i^T\boldsymbol{\theta})^2\right) c(\nu)\lambda_i^{\frac{\nu}{2}-1}\exp\left(\frac{-\nu\lambda_i}{2}\right)\right\}, \quad (3.75)\end{aligned}$$

where $c(\nu) = \left[\Gamma(\frac{\nu}{2})(\frac{\nu}{2})^{(\nu/2)}\right]^{-1}$ and $\pi(\nu)$ is the prior distribution on ν. The fully conditional distribution of Z_i are independent with

$$\begin{cases} \text{truncated at the left by} & 0 \text{ if } y_i = 1 \\ \text{truncated at the right by} & 0 \text{ if } y_i = 0 \end{cases} \quad (3.76)$$

The conditional distributions of $\boldsymbol{\theta}$, λ and ν are given by

$$(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{\lambda}, \nu) \sim N_k(\hat{\boldsymbol{\theta}}_{\boldsymbol{Z},\boldsymbol{\lambda}}, (\acute{X}WX)^{-1}), \quad (3.77)$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{Z},\boldsymbol{\lambda}} = (\acute{X}WX)^{-1}\acute{X}WZ$ and $W = diag(\lambda_i)$,

$$(\lambda|\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{\theta}, \nu) \sim Gamma\left(\frac{\nu+1}{2}, \frac{2}{\nu+(Z_i - \boldsymbol{x}_i^T\boldsymbol{\theta})^2}\right), \tag{3.78}$$

and $(\nu|\boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda})$ is distributed with a pdf proportional to

$$\pi(\nu)\prod_{i=1}^{N}(c(\nu)\lambda_i^{\nu/2-1}\exp(\frac{-\nu\lambda_i}{2})). \tag{3.79}$$

To implement the Gibbs sampler, Albert and Chib (1993) started with $\boldsymbol{\theta}$ equal to the least squares estimate under the probit model and set $\lambda_i = 1 \quad \forall i$. Parameter values are then sampled in the following order: Equations (3.76), (3.77), (3.78) and (3.79).

Halekoh et al. (2004) uses discrete choice experiments to assess the choices made by an animal. The model used is similar to those used to model discrete choices between health states. The example used in Halekoh et al. (2004) involves pigs and rooting material. Two categories of rooting material were used, containing three different rooting materials. For each category the rooting materials were each placed in one arm of a three arm maze. Two pigs were lead to the centre of the maze and after one and a half minutes the position of the pigs determined their choice. If both pigs were in one arm of the maze, the rooting material in that arm was given as their choice. If both pigs remained in the centre of the maze, they had no preference. It was not possible to conduct the experiment where the rooting material could be presented independently from the orientation of the maze arms. As the effect of the orientation of the maze arms could not be excluded, three combinations of the orientation of the maze arms and the rooting materials were tested with different pigs. Twelve pairs of pigs had their choices recorded on four occasions for the same option of maze arm and material

combination.

A multinomial logistic random intercepts model was fitted to the data. If pair of pigs i $(i = 1, ...12)$, chooses one of k $(k = 1, 2, 3, 4)$ options on occasion t $(t = 1, 2, 3, 4)$, let Y_{it} be the random variable describing the choice of animal i at time t, i.e. $Y_{it} = k$ if option k is chosen. Option k is equal to 1, 2, or 3 if one of the rooting materials are chosen, and equal to 4 if no choice is made.

The probability of animal i choosing option k on occasion t is given by

$$P(Y_{it} = k|\sigma^2, \epsilon_{ik}, x_{itk}) = \theta_{itk} = \frac{\exp \mu_{itk}}{\sum_{i=1}^{K} \exp \mu_{itk}} \tag{3.80}$$

where

$$\mu_{itk} = \alpha_k + \boldsymbol{v}'_{c(i)k}\boldsymbol{\delta} + \epsilon_{ik} \quad \text{for} \quad k < 4 \tag{3.81}$$

$$\mu_{it4} = \epsilon_{i4}. \tag{3.82}$$

The parameter $\boldsymbol{v}'_{ck}$ is row k of the design matrix V_c for combination c of the options of rooting materials and maze arms. There are three options for V_c:

$$V_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; V_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{ and } V_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The parameters $\boldsymbol{\delta} = (\delta_l, \delta_s, \delta_r)$ represented the orientation effect.

The ϵ_{ik} are random intercepts allowing for animal specific choice probabilities and are assumed to be independently normally distributed: $\epsilon_{ik}|\sigma^2 \sim N(0, \sigma^2)$. The vector of random intercepts for animal i is denoted by $\epsilon_i = (\epsilon_{i1}, .., \epsilon_{i4})$. The components of the parameter vector $\boldsymbol{\beta}$ and the variance σ^2 were given non-informative prior distributions. Each parameter in $\boldsymbol{\beta}$ was given a $N(0, 10^6)$

prior distribution and the standard deviation σ had a uniform prior distribution, $U(0, 100)$. Assuming independence between the prior distributions, the posterior distribution is given by

$$f(\beta, \sigma^2, \epsilon_1, ...\epsilon_1 2|Y) \propto \prod_{i=1}^{12} \prod_{k=1}^{4} \prod_{t=1}^{4} \theta_{itk}^{y_{itk}} \varphi(\epsilon_{ik}/\sigma) f(\beta) f(\sigma^2) \tag{3.83}$$

The vector $Y = (y_{itk})$ denoted all the responses where $y_{itk} = 1$ if animal i chooses option k on occasion y.

The posterior distribution of the choice probabilities can be used to describe a future choice of a new animal, which is assumed to be sampled from the same population as the animals used in the experiment. The vector of choice probabilities for some new animal i^* is denoted by $\theta(\psi, X_t)$, where $\psi = (\beta, \sigma^2, \epsilon_{i^*})$ and X_t contains the vectors x'_{i^*tk} containing variables relating to eh choice k made by animal i^* at time t. The choice probabilities for the new animal are then predicted by the mean:

$$\boldsymbol{\theta}(\boldsymbol{X}_t) = \int \boldsymbol{\theta}(\boldsymbol{\psi}|\boldsymbol{X}_t) f(\boldsymbol{\psi}|\boldsymbol{Y}) d\boldsymbol{\psi}$$

where $f(\psi|Y) = \prod_{k=1}^{K} \varphi(\epsilon_{i^*k}/\sigma) f(\beta, \sigma^2|Y)$

The probabilities $\theta(X_t)$ can be interpreted as the choice probabilities for new responses Y_{i^*k}.

Discrete choice experiments are used extensively in both transport and environmental research. A search of articles using Bayesian approaches in discrete choice experiments are limited. However, Lemp et al. (2010) provides an example where Bayesian methods are used. The continuous cross-nested logit model is introduced and an application of the model to work-tour departure times was

estimated using Bayesian techniques.

3.10 Summary

In this chapter we have reviewed discrete choice experiments and explained how the resulting discrete choice data can be modelled to estimate health state utilities. Three models are reviewed: logit, probit and mixed logit, and properties of each model are discussed, including identifiability of additive constants and identifiability of error variances. To include constants in discrete choice models, one constant must be normalised to zero and all other constants estimated relative to the normalised constant. A multiplicative constant will increase the utility scale and the variance of the errors. To identify values for a multiplicative constant the utility scale must be normalised. The standard way to do this is to fix the value of the error variance term to 1. However, it can also be achieved by fixing the utilities of perfect health and death. The property of independence of alternatives is also defined. The logit model satisfies this property whilst the probit and mixed logit do not.

Chapter 4

Analysis of AQL-5D data

4.1 Introduction

The AQL-5D classification system defines health states specific to asthma. There are five attributes in the classification system where each attribute has five levels of severity. It is described further in Section 2.5.4. This chapter concerns the use of discrete choice data to estimate utilities defined the AQL-5D classification system. Logit and probit models are fitted to the data, first using maximum likelihood. A Bayesian approach is then considered for both the logit and probit models. Posterior distributions are generated for the parameters in the two model using MCMC sampling given the three priors: Gamma(1,10), Gamma(5,15) and Uniform(0,1).

4.2 Data

The data used are pairwise choice data from an existing study designed to show how rank and discrete choice data can be used to generate utilities for health states defined by the AQL-5D classification system. Brazier et al. (2006) present an analysis of these data, where a probit model is fitted to the data. In this chap-

ter uncertainty in the utility estimates is investigated by considering a Bayesian approach, which are compared to estimates derived using maximum likelihood.

A sample of 307 people (40% response rate) in South Yorkshire were interviewed to elicit preferences for a sample of AQL-5D health states. A balanced design was used to select 98 of the possible 3124 health states. The total score of the levels of each attribute was calculated for each health state and this score was used to stratify the 98 health states into severity groups. Health states from each group were randomly allocated into 14 blocks, so that each block contained 7 health states. Each respondent interviewed was allocated one block.

In the interview the first task was to rank 10 health states in order of preference. The health states considered were perfect health, immediate death, the worst health state defined by the AQL-5D classification system and the 7 health states in the block allocated to the respondent. The next task was to elicit Time Trade-off values for 8 health states; these were the 7 health states in the respondent's block and the worst AQL-5D health state.

Following the interview each respondent that consented was sent a discrete choice experiment questionnaire. An application in SAS developed by Huber and Zwerina (1996) was used to select the health states for the discrete choice experiment. The application obtains an optimal statistical design based on level balance, orthogonality, minimal overlap and utility balance. The SAS program selected 24 pairwise comparisons from the AQL-5D classification system. Each pairwise comparison was randomly allocated to one of four versions of the questionnaire, so that each questionnaire contained 6 of the selected pairwise comparisons. Two additional pairwise comparisons were included in each questionnaire, comparing immediate death with an AQL-5D health state. Health states in the AQL-5D that could be regarded as worse than death by some respondents were chosen to be compared

to immediate death.

In addition to the pairwise comparisons, each questionnaire was used to ask questions about the respondent's background, general health, previous experience of asthma, whether or not they currently have asthma, and their description of asthma using the AQL-5D classification system. Each respondent was sent one questionnaire.

Of the 307 people interviewed, 168 returned a completed questionnaire, generating 1336 observed pairwise comparisons. Table 4.1 summarises characteristics of the people in the sample. Table 4.2 summarises the numbers of respondents with asthma at the time of completing the questionnaire and the numbers with experience of asthma respectively.

		Number	Percentage
Sex	Male	72	43%
	Female	96	57%
Age	18-25	6	3.6%
	26-35	22	13.1%
	36-45	28	16.7%
	46-55	38	22.6%
	55-65	39	23.2%
	over 65	35	20.8%

Table 4.1: Number and percentage of respondents by age and sex

		Number	Percentage
Current Asthma	Yes	35	21%
	No	131	78%
	Unknown	2	1%
Current or Previous Asthma	Yes	44	26%
	No	110	66%
	Unknown	14	8%

Table 4.2: Number and percentage of respondents with experience of Asthma

The design used for this discrete choice experiment is described by Huber and Zwerina (1996). Some other articles where discrete choice design is reviewed are Kessels et al. (2004), Kessels et al. (2006), Kuhfeld et al. (1994) and Viney et al. (2005).

4.3 Modelling discrete choices

We wish to obtain the population mean utility for every health state defined by the AQL-5D classification system. A parametric model is fitted to the data with the parameters defining the population mean utility of any health state in the AQL-5D classification system.

In Chapter 3, $\boldsymbol{x}$ is defined to be a vector of dummy variables that defines an alternative. In this chapter, let $\boldsymbol{x} = (x_1, x_2, ..., x_{21})$ be a vector of 21 dummy variables that defines either a health state in the AQL-5D classification system or the health state death. For $a = 1, ...5$ and $b = 1, ...4$ each element of $\boldsymbol{x}$ is defined as

$$x_{4(a-1)+b} = \begin{cases} 1 & \text{if attribute } a \text{ is at level } b+1 \text{ or higher} \\ 0 & \text{if attribute } a \text{ is at level 1} \end{cases} \tag{4.1}$$

The element x_{21} is zero unless $\boldsymbol{x}$ represents the health state death, in which case

$$x_k = \begin{cases} 0 & k = 1, ...20, \\ 1 & k = 21. \end{cases} \tag{4.2}$$

Define $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$ to be a pair of health states defined by the AQL-5D classification system that are assessed by individual i. Define U_{ij} to be the utility individual i has for health state $\boldsymbol{x}_{ij}$. The relationship between U_{ij} and $\boldsymbol{x}_{ij}$ as presented in equation (3.2) is $U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}$. In the discrete choice experiment each individual assesses eight comparisons, two of which include the health state death. By definition, all individuals have a utility of zero for the health state death. For these comparisons assume $\boldsymbol{x}_{i2}$ is the health state death. Therefore $g(\boldsymbol{x}_{i2}) = 0$ and $\varepsilon_{i2} = 0$.

In a discrete choice experiment an individual i selects the preferred health state in set B. The probability that individual i chooses health state $\boldsymbol{x}_{i1}$ is equivalent to the probability that the utility individual i has for health state $\boldsymbol{x}_{i1}$ is greater than the utility individual i has for health state $\boldsymbol{x}_{i2}$. Define $P_B(\boldsymbol{x}_{i1})$ to be the probability that individual i will prefer health state $\boldsymbol{x}_{i1}$ to health state $\boldsymbol{x}_{i2}$. Using equation (3.3) the probability can be written as

$$P_B(x_{i1}) = P[g(\boldsymbol{x}_{i2}) + \varepsilon_{i2} < g(\boldsymbol{x}_{i1}) + \varepsilon_{i1}] \tag{4.3}$$

If $\boldsymbol{x}_{i2}$ is the health state death

$$P_B(x_{i1}) = P[0 < g(\boldsymbol{x}_{i1}) + \varepsilon_{i1}]. \tag{4.4}$$

A linear model is assumed for the mean health state utility and $g(\boldsymbol{x}_{ij}), j = 1, 2$, is defined to be $g(\boldsymbol{x}_{ij}) = 1 - \boldsymbol{x}_{ij}^T \boldsymbol{\theta}$. The vector $\boldsymbol{\theta}$ contains 21 unknown parameters $\boldsymbol{\theta} = (\theta_1, ...\theta_{20}, \theta_{21})^T$. Each of the first 20 parameters represent the decrease in

utility associated with the increase of an attribute by one level. The parameter θ_{21} represents the decrease in utility from perfect health to death. If $\boldsymbol{x}_{i2}$ is the health state death, $\boldsymbol{x}_{i2}^T\boldsymbol{\theta} = 1$ and $U_{i2} = 0$. Therefore $\theta_{21} = 1$.

Two models are considered for the data, the logit and probit models.

4.3.1 LOGIT MODEL

As shown in Section 3.5, if the errors ε_{ij} are assumed to have a type 1 extreme value distribution then the probability associated with choosing a health state is given by the logit model. The mean of the errors is defined to be zero and therefore the location parameter is $\mu = -0.5722\sigma$, where σ is the scale parameter of the distribution. The pdf is given by equation (3.25). Using equation (3.25) and equation (3.28), it can be shown that the probability that individual i prefers health state $\boldsymbol{x}_{i1}$ to health state $\boldsymbol{x}_{i2}$ is

$$P_B(\boldsymbol{x}_{i1}) = \frac{\exp\left(\frac{g(\boldsymbol{x}_{i1})}{\sigma}\right)}{\exp\left(\frac{g(\boldsymbol{x}_{i1})}{\sigma}\right) + \exp\left(\frac{g(\boldsymbol{x}_{i2})}{\sigma}\right)}. \tag{4.5}$$

Some comparisons include the health state death. If $\boldsymbol{x}_{i2}$ is the health state death, then the probability of choosing health state $\boldsymbol{x}_{i1}$ is

$$P_B(\boldsymbol{x}_{i1}) = 1 - \exp\left(-\exp\left(\frac{g(\boldsymbol{x}_{i1}) - 0.5722\sigma}{\sigma}\right)\right). \tag{4.6}$$

Each respondent in the study assesses 8 pairwise comparisons. Define $B_{ij} = \{\boldsymbol{x}_{ij1}, \boldsymbol{x}_{ij2}\}$ to be the set containing the j^{th} pair of health states to be compared by individual i. The sets compared by individual i can be written as $B_{i1} = \{\boldsymbol{x}_{i11}, \boldsymbol{x}_{i12}\}, B_{i2} = \{\boldsymbol{x}_{i21}, \boldsymbol{x}_{i22}\}, ..., B_{i8} = \{\boldsymbol{x}_{i81}, \boldsymbol{x}_{i82}\}$. If the j^{th} comparison

made by individual i is between health state $\boldsymbol{x}_{ij1}$ and $\boldsymbol{x}_{ij2}$, define

$$n_{ij} = \begin{cases} 1 & \text{if individual } i \text{ prefers } \boldsymbol{x}_{ij1} \text{ to } \boldsymbol{x}_{ij2}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

For comparisons that include the health state death, assume $\boldsymbol{x}_{ij2}$ is the health state death. Define $M = \{m_{ij}, i = 1, ..., I, j = 1, ..., 8\}$ with

$$m_{ij} = \begin{cases} 0 & \text{if health state } x_{ij2} \text{ is death,} \\ 1 & \text{otherwise.} \end{cases} \tag{4.8}$$

and $N = \{n_{ij}, i = 1, ..., I, j = 1, ...8\}$. The likelihood function for the individuals' choices is

$$\begin{aligned} P(N|\boldsymbol{\theta}, \sigma^2, M) = \prod_{i=1}^{I} \prod_{j=1}^{8} \Bigg\{ & \left[\frac{\exp\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right)}{\exp\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right) + \exp\left(\frac{g(\boldsymbol{x}_{ij2})}{\sigma}\right)} \right]^{n_{ij} \times (1 - m_{ij})} \\ & \times \left[1 - \frac{\exp\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right)}{\exp\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right) + \exp\left(\frac{g(\boldsymbol{x}_{ij2})}{\sigma}\right)} \right]^{(1 - n_{ij})(1 - m_{ij})} \\ & \times \left[1 - \exp\left(- \exp\left(\frac{g(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right]^{n_{ij} \times m_{ij}} \\ & \times \left[\exp\left(- \exp\left(\frac{g(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right]^{(1 - n_{ij}) \times m_{ij}} \Bigg\}, \end{aligned} \tag{4.9}$$

with $g(\boldsymbol{x}_{ijk}) = 1 - \boldsymbol{\theta}^T \boldsymbol{x}_{ijk}$. The log likelihood is

$$\log P(N|\boldsymbol{\theta}, \sigma^2, M) = \sum_{i=1}^{I} \sum_{j=1}^{8} \left\{ n_{ij}(1 - m_{ij}) \left[\frac{g(\boldsymbol{x}_{ij1}) - g(\boldsymbol{x}_{ij2})}{\sigma} \right] + (1 - m_{ij}) \frac{g(\boldsymbol{x}_{ij2})}{\sigma} \right.$$

$$- (1 - m_{ij}) \log \left[\exp\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma} \right) + \exp\left(\frac{g(\boldsymbol{x}_{ij2})}{\sigma} \right) \right]$$

$$+ (n_{ij} \times m_{ij}) \times \log \left[1 - \exp\left(- \exp\left(\frac{g(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right]$$

$$\left. + (1 - n_{ij}) \times m_{ij} \times \left(- \exp\left(\frac{g(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right\}.$$

4.3.2 Probit Model

If an individual i has a set of possible health states $B = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$, then from equation (4.4) the probability of choosing one of the pair of alternatives, $P_B(\boldsymbol{x}_{i1})$, can be written as

$$P_B(\boldsymbol{x}_{i1}) = P[\varepsilon_{i2} - \varepsilon_{i1} < g(\boldsymbol{x}_{i1}) - g(\boldsymbol{x}_{i2})]. \tag{4.10}$$

If the errors ε_{i1} and ε_{i2} are assumed to be independent and identically distributed with normal distributions,

$$\varepsilon_{i1} \sim N(0, \sigma^2), \qquad \varepsilon_{i2} \sim N(0, \sigma^2), \tag{4.11}$$

then the difference $\varepsilon = \varepsilon_{i2} - \varepsilon_{i1}$ also has a normal distribution, $\varepsilon \sim N(0, 2\sigma^2)$. Therefore, the probability $P_B(\boldsymbol{x}_{i1})$ is given by

$$P_B(\boldsymbol{x}_{i1}) = \Phi\left(\frac{g(\boldsymbol{x}_{i1}) - g(\boldsymbol{x}_{i2})}{\sqrt{2\sigma^2}} \right), \tag{4.12}$$

where Φ is the standard normal cdf. Equation (4.12) is called the probit model. If $\boldsymbol{x}_{i2}$ is the health state death then the probability is written as

$$P_B(\boldsymbol{x}_{i1}) = P[\varepsilon_{i1} > -g(\boldsymbol{x}_{i1})]. \tag{4.13}$$

If ε_{i1} is assumed to have a normal distribution, $\varepsilon_{i1} \sim N(0, \sigma^2)$, the probability, $P_B(\boldsymbol{x}_{i1})$ is given by

$$P_B(\boldsymbol{x}_{i1}) = \Phi\left(\frac{g(\boldsymbol{x}_{i1})}{\sigma}\right), \tag{4.14}$$

where Φ is the standard normal cdf.

To fit the probit model, values need to be inferred for $\boldsymbol{\theta}$. If an individual i has 8 pairs of health states to compare, $B_{ij} = \{\boldsymbol{x}_{ij1}, \boldsymbol{x}_{ij2}\}$, $j = 1, ..., 8$, and n_{ij}, m_{ij}, M and N are as defined in section 4.3.1, the likelihood of the individuals' choices is

$$\begin{aligned} P(N|\boldsymbol{\theta}, \sigma^2, M) = \prod_{i=1}^{I}\prod_{j=1}^{8} \Bigg\{ & \left[\Phi\left(\frac{g(\boldsymbol{x}_{ij1}) - g(\boldsymbol{x}_{ij2})}{\sqrt{2\sigma^2}}\right)\right]^{n_{ij} \times (1 - m_{ij})} \\ & \times \left[1 - \Phi\left(\frac{g(\boldsymbol{x}_{ij1}) - g(\boldsymbol{x}_{ij2})}{\sqrt{\sigma^2}}\right)\right]^{(1 - n_{ij}) \times (1 - m_{ij})} \\ & \times \left[\Phi\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right)\right]^{n_{ij} \times m_{ij}} \times \left[1 - \Phi\left(\frac{g(\boldsymbol{x}_{ij1})}{\sigma}\right)\right]^{(1 - n_{ij}) \times m_{ij}} \Bigg\}, \end{aligned} \tag{4.15}$$

again with $g(\boldsymbol{x}_{ijk}) = 1 - \boldsymbol{\theta}^T \boldsymbol{x}_{ijk}$. The log likelihood is

$$\log P(N|\boldsymbol{\theta}, \sigma^2, M) = \sum_{i=1}^{I} \sum_{j=1}^{8} \left\{ n_{ij} \log \Phi \left(\frac{g(\boldsymbol{x}_{ij1}) - g(\boldsymbol{x}_{ij2})}{\sqrt{2\sigma^2}} \right) \right.$$

$$+ (1 - n_{ij}) \log \left[1 - \Phi \left(\frac{g(\boldsymbol{x}_{ij1}) - g(\boldsymbol{x}_{ij2})}{\sqrt{2\sigma^2}} \right) \right]$$

$$+ (n_{ij} \times m_{ij}) \times \log \left[\Phi \left(\frac{g(\boldsymbol{x}_{i1})}{\sigma} \right) \right]$$

$$\left. (1 - n_{ij}) \times m_{ij} \times \log \left[1 - \Phi \left(\frac{g(\boldsymbol{x}_{i1})}{\sigma} \right) \right] \right\}.$$

4.4 Maximum Likelihood Results

4.4.1 Logit Model

The logit model described in section 3.5 is fitted to the AQL-5D data using maximum likelihood.

Attribute	Concern	Breath	Weather	Sleep	Activities
Level 2	0.0012	0.0045	8.9261×10^{-6}	6.5952×10^{-7}	0.0316
Level 3	8.697×10^{-8}	1.9007×10^{-7}	0.0221	0.0565	0.0268
Level 4	0.1350	0.1248	0.0795	0.0223	0.2421
Level 5	0.0334	5.0461×10^{-7}	0.0256	0.0287	0.0003
Scale			0.2419		

Table 4.3: Maximum Likelihood estimates for a logit model (rounded to 4 d.p)

Table 4.3 shows the parameter estimates after fitting the logit model to the AQL-5D data using maximum likelihood. Each parameter estimate represents the incremental decrease in mean utility when an attribute increases by one level of severity. For example, the estimated mean decrease in utility when changing

from the attribute Activities at level 1 (not at all limited in any activity done) to level 4 (very limited in every activity done) is $0.0316+0.0268+0.2421 = 0.30205$. As the level of the attribute increases the size of the parameter estimate does not necessarily increase. Several parameter estimates have very small values and are 0 when rounded to four decimal places. For example, the decrease in mean utility when the attribute Weather increases from level 1 to level 2 is 8.9261×10^{-6}. This implies that the health state $(1, 1, 2, 1, 1)$ is almost identical to perfect health in mean utility. For most attributes the largest decrease in mean utility is a change from level 3 to level 4. Arguably, for each attribute, the description of level 3 is nearer to level 2 than level 4 and the description of level 4 is nearer to level 5 than level 3. The largest decrease in mean utility is a change from level 3 to level 4 for the attribute Activities. This decreases the mean utility by 0.2421. The attribute Concern accounts for a larger decrease in mean utility than the attribute Breath, which would probably not be expected.

If all attributes in a health state are at level 5 the health state is regarded as the worst health state described by the AQL-5D classification system. It is important to find the estimated mean utility of the worst health state. This shows the complete range of the mean utilities and whether any health states are considered worse than death. The decrease in mean utility from perfect health to the worst health state is the sum of all the parameter estimates and is equal to 0.8347. The mean utility of the worst health state is therefore 0.1653 and so is considered to be better than death.

4.4.2 Probit Model

The probit model is fitted to the AQL-5D data using maximum likelihood estimation. Table 4.4 shows the parameter estimates after fitting the probit model. The decrease in mean utility from perfect health to the worst health state is 0.8641.

The mean utility of the worst health state is therefore 0.1359, which is smaller than in the logit model.

Attribute	Concern	Breath	Weather	Sleep	Activities
Level 2	0.0045	0.0001	0.0001	0.0093	0.0446
Level 3	0.0007	0.0189	0.0449	0.0561	0.0341
Level 4	0.1084	0.1153	0.0514	0.0172	0.1902
Level 5	0.0223	0.0002	0.0546	0.0431	0.0483
Scale			0.2307		

Table 4.4: Maximum Likelihood estimates for a probit model (rounded to 4 d.p)

Mean utilities are calculated for 48 health states in the AQL-5D classification system, using the parameter estimates from both the logit model and the probit model. Figure 4.1 presents a graph of the mean utilities calculated using estimates from both models, plotted in decreasing order of the utility from the logit model. The health states tend to have larger estimated utilities in the logit model.

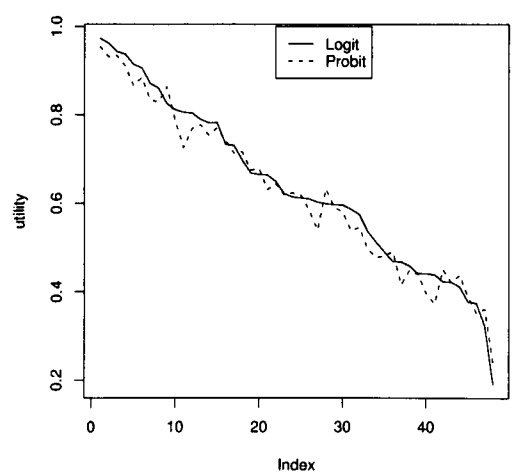


Figure 4.1: Mean utilities for 48 health states using maximum likelihood estimates for the probit and logit models

4.5 Bayesian inference for health state utilities

Using a classical approach parameter estimates, such as those estimated using maximum likelihood, are used to calculate one estimated utility for every health state in the AQL-5D classification system. The estimated utilities can then be used to represent QALYs in an economic evaluation. A Bayesian approach considers the uncertainty present in each utility measurement. A posterior distribution is derived for each parameter and therefore a posterior distribution can also be derived for the utility of each health state. This represents the range of possible values for the utility of each health state. When uncertainty is considered in an economic evaluation we can then consider a number of possible QALYs for each health state rather than just one.

Bayesian inference derives a probability distribution of the set of unknown parameter $\boldsymbol{\pi}$. It can also incorporate prior beliefs about the parameter $\boldsymbol{\pi}$. The prior distribution, $P(\boldsymbol{\pi})$, is the probability distribution of parameter $\boldsymbol{\pi}$ before the data N are observed. The posterior distribution, which is the probability distribution of the parameter $\boldsymbol{\pi}$ after the data are observed, is calculated using

$$P(\boldsymbol{\pi}|N) = \frac{P(N|\boldsymbol{\pi})P(\boldsymbol{\pi})}{P(N)}. \tag{4.16}$$

4.5.1 Prior distribution

Prior distributions are needed for the set of parameters $\boldsymbol{\pi} = (\boldsymbol{\theta}, \sigma)$ where $\boldsymbol{\theta} = \{\theta_1, ...\theta_{21}\}$ and σ are defined in section 4.3. We first consider $\boldsymbol{\theta}$. As defined previously $\theta_{21} = 1$. The parameters $\theta_1, ...\theta_{20}$ each represent the decrease in utility associated with the increase of an attribute by one level. As no health states can have a utility greater than the utility of perfect health, the parameter values must be positive. Consequently we first consider independent gamma prior

distributions for the elements of $\boldsymbol{\theta}$. The pdf of the gamma distribution is given by

$$f(\theta_i) = \frac{\beta^\alpha}{(\alpha-1)!}\theta^{\alpha-1}e^{-\beta\theta}, i = 1, ...20, \tag{4.17}$$

where α is the shape parameter and β is the rate parameter. The mean of the gamma distribution is $\frac{\alpha}{\beta}$ and the variance is $\frac{\alpha}{\beta^2}$. Several values of α and β are considered. A single parameter value is not likely to be greater than 1 as a change in the level of one attribute would not be expected to produce a change in utility greater than the change from perfect health to death. Therefore Gamma distributions with mass concentrated in the range (0,1) are considered.

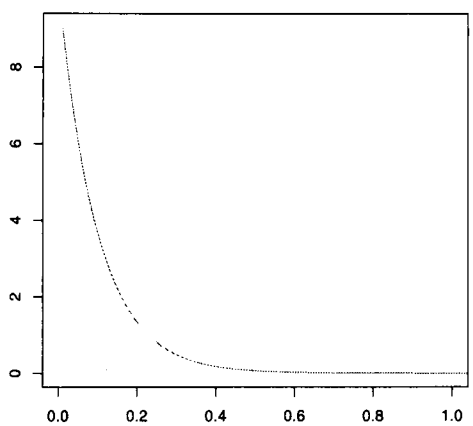


Figure 4.2: Gamma(1,10) Prior

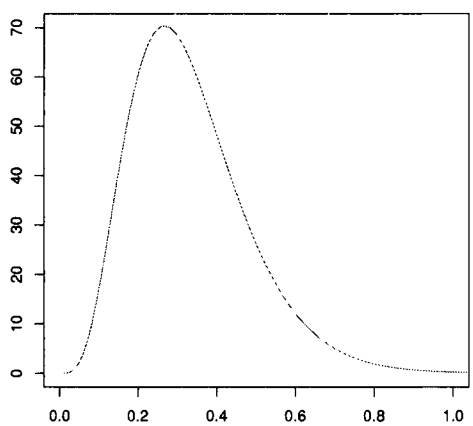


Figure 4.3: Gamma(5,15) Prior

Two gamma prior distributions are considered, a Gamma(1,10) distribution and a Gamma(5,15) distribution. Figure 4.2 shows the probability distribution of the Gamma(1,10) distribution, which has scale parameter $\alpha = 1$ and rate parameter $\beta = 10$. This prior assumes that the parameters are more likely to be closer to 0 and have a small probability of having a value greater than 0.4. Figure 4.3 shows the probability distribution of the Gamma(5,15) distribution, which has scale parameter $\alpha = 5$ and rate parameter $\beta = 15$. This prior distribution assumes

that the parameter values are unlikely to be close to 0 and have a large probability of being between 0.2 and 0.4. A uniform prior distribution over the range (0,1) is also used. This prior distribution assumes that the parameter values are equally likely to be any value between 0 and 1. The parameter σ is related to the variance of the distribution assumed for the errors, ε_{ij}, in equation (3.2). When fitting a model, the variance of the errors is required to be as small as possible. To explain this, consider an example where the errors are assumed to have a normal distribution $\varepsilon \sim N(0, \sigma^2)$ with variance $\sigma^2 = 1$. The distribution of the errors is given by figure 4.4. Most of the distribution is between $\mu - 2\sigma = -2$ and $\mu + 2\sigma = \mu + 2$. Suppose the population mean utility of an AQL-5D health state is $g(\boldsymbol{x}_{ij}) = 0.8$. This error distribution would imply that the utilities for this health state can vary between -1.4 and 2.8. This is unrealistic for two reasons. This means the utility scale for that health state is four times the size of the scale between death and perfect health, and no health state can have a utility greater than 1. Alternatively suppose the variance is small, for example, $\sigma^2 = 0.05^2 = 0.0025$. The utility of any given health state will vary by between -0.01 and $+0.01$. Therefore, as with the parameters $\boldsymbol{\theta}$, σ is assumed to be less than 1. We start by trying the same prior as used for the parameters $\boldsymbol{\theta}$ and will then consider other priors.

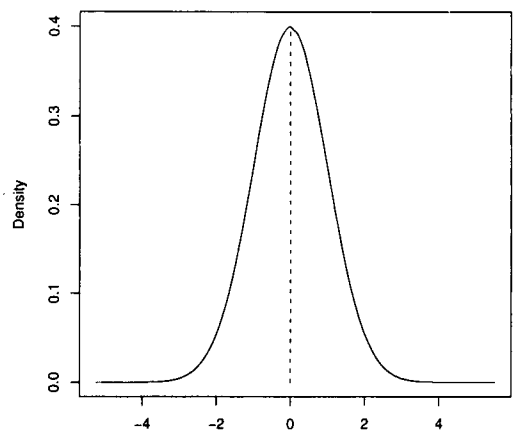


Figure 4.4: Normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$

4.5.2 Posterior Inference using MCMC

Given the likelihood and prior distributions considered in this chapter, the posterior distributions cannot be derived analytically. Therefore MCMC sampling is used to simulate values from the posterior distribution. The posterior distributions of the 21 unknown parameters in the probit and logit models for the AQL-5D data can be inferred using simulation. MCMC simulation samples vectors of parameter values, $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_{21})$. For notational ease we write $\pi_i = \theta_i \quad i = 1, ...20$ and $\pi_{21} = \sigma$. The sequence of vectors $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, ...$ is a Markov Chain. The stationary distribution of the Markov chain is the required posterior distribution.

Let the state of the Markov chain at time t be $\boldsymbol{\pi}^{(t)} = (\pi_1^{(t)}, ..., \pi_{21}^{(t)})$. To update this vector to the state at time $t+1$, $\boldsymbol{\pi}^{(t+1)} = (\pi_1^{(t+1)}, ..., \pi_{21}^{(t+1)})$, a new value is generated for each of the 21 parameters. The process of updating the vector $\boldsymbol{\pi}^{(t)}$ to the new vector $\boldsymbol{\pi}^{(t+1)}$ is referred to as one iteration of the MCMC.

If $\boldsymbol{\pi}_{-i}$ is defined as the vector of all elements of $\boldsymbol{\pi}$ except π_i,

$$\boldsymbol{\pi}_{-i} = (\pi_1, .., \pi_{i-1}, \pi_{i+1}, ..., \pi_{21}), \tag{4.18}$$

$\boldsymbol{\pi}_{-i}^{(t)}$ can be defined as the state of $\boldsymbol{\pi}_{-i}$ after updating the element π_{i-1} at time $t+1$,

$$\boldsymbol{\pi}_{-i}^{(t)} = (\pi_1^{(t+1)}, ..., \pi_{i-1}^{(t+1)}, \pi_{i+1}^{(t)}, ...\pi_{21}^{(t)}). \tag{4.19}$$

The method of MCMC used is the Metropolis-Hastings algorithm. This is described as follows. To start the MCMC a vector of starting values is needed, $\boldsymbol{\pi}_1 = (\pi_{1,1}....\pi_{1,21})$. At each step within an iteration of the MCMC, a candidate value, π, is sampled from the proposal distribution $q_i(\pi|\pi_i^{(t)})$, where $\pi_i^{(t)}$ is the current value of the parameter π_i. We have already noted that each parameter must be positive and so this must be considered when choosing the proposal distribution. A lognormally distributed variable must be positive and so this distribution is considered as the proposal distribution,

$$q_i(\pi|\pi_i^{(t)}) = \frac{1}{\pi\nu\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log\pi - \log\pi_i^{(t)}}{\nu}\right)^2\right], \tag{4.20}$$

where ν is the standard deviation of $\log(\pi)$. A value of ν is chosen for each parameter to allow between 40% and 60% of samples generated to be accepted.

After each candidate value is sampled the acceptance probability is calculated, which is used to decide whether to accept the candidate value for a parameter. If the current state of the Markov chain is $(\boldsymbol{\pi}_{-i}^{(t)}, \pi_i^{(t)})$, the acceptance probability for a candidate value π is

$$\alpha(\boldsymbol{\pi}_{-i}^{(t)}, \pi_i^{(t)}, \pi) = min\left(1, \frac{f(\pi)f(N|\boldsymbol{\pi}_{-i}^{(t)}, \pi)q_i(\pi_i^{(t)}|\pi)}{f(\pi_i^{(t)})f(N|\boldsymbol{\pi}_{-i}^{(t)}, \pi_i^{(t)})q_i(\pi|\pi_i^{(t)})}\right), \tag{4.21}$$

where $f(\pi)$ and $f(\pi_{t,i})$ are the prior distributions for the candidate value and current value of parameter π_i respectively, and $f(N|\boldsymbol{\pi}_{-i}^{(t)}, \pi)$ and $f(N|\boldsymbol{\pi}_{-i}^{(t)}, \pi_i^{(t)})$ are the corresponding likelihoods. To decide whether to accept the candidate value π, a random number, u, is generated from the uniform distribution $U[0,1]$. If $u < \alpha(\boldsymbol{\pi}_{-i}^{(t)}, \pi_i^{(t)}, \pi)$ then $\pi_i^{(t+1)} = \pi$; otherwise $\pi_i^{(t+1)} = \pi_i^{(t)}$.

The iterations are continued until the Markov chain converges to the stationary distribution. If this occurs at time T then $(\pi_i^{(T+1)}, ...\pi_i^{(T+n)})$ will be a sample from the posterior distribution for parameter π_i. The sample is used to calculate the posterior mean, $E[\pi_i|N]$, and 95% posterior intervals for each parameter.

To assess convergence the parameter values are plotted against the iteration number. A second sample can also be selected from the MCMC output, $(\pi_i^{(T+n+1)}, ...\pi_i^{(T+2n)})$. If the distribution has converged to the stationary distribution, the mean and 95% posterior intervals of the second sample should be similar to those calculated using the first sample.

4.5.3 Residuals

Residuals are used to examine the adequacy of a model in predicting individual data points. Outliers can affect parameter estimates, increase the estimates of variance parameters and often indicate a deficiency in the model.

There are 32 sets of pairwise choices in the AQL-5D data. Each set can be written as $B_j = \{\boldsymbol{x}_{j1}, \boldsymbol{x}_{j2}\}$. Suppose there are K vectors of parameters sampled from the MCMC, $\boldsymbol{\pi}_k, k = 1, ...K$. A Bayesian residual is defined in Johnson and Albert

(1999) as

$$r_{j,k} = \frac{n_j}{N_j} - P_{j,k}, \quad j = 1, ...32, \quad k = 1, ..., K, \tag{4.22}$$

where N_j is the number of individuals assessing set B_j, n_j is the number of individuals preferring health state $\boldsymbol{x}_{j1}$ to health state $\boldsymbol{x}_{j2}$, and $P_{j,k}$ is the fitted proportion of individuals preferring health state $\boldsymbol{x}_{ij1}$ to health state $\boldsymbol{x}_{ij2}$. The fitted proportion, $P_{j,k}$ is calculated using each parameter vector $\boldsymbol{\pi}_k$. For the logit model the fitted proportion is

$$P_{j,k} = \frac{\exp\left(\frac{\boldsymbol{x}_{j1}^T \boldsymbol{\theta}_k}{\sigma}\right)}{\exp\left(\frac{\boldsymbol{x}_{j1}^T \boldsymbol{\theta}_k}{\sigma}\right) + \exp\left(\frac{\boldsymbol{x}_{j2}^T \boldsymbol{\theta}_k}{\sigma}\right)}, \tag{4.23}$$

If $\boldsymbol{x}_{j2}$ is the health state death, then $P_{j,k}$ is

$$P_{j,k} = \left(1 - \exp\left(-\exp\left(\frac{\boldsymbol{x}_{j1}^T \boldsymbol{\theta}_k - 0.5722\sigma}{\sigma}\right)\right)\right). \tag{4.24}$$

Similar equations can be derived for the probit model. The mean of the fitted probabilities is also calculated for each pairwise comparison by

$$P_j = \frac{\sum_{k=1}^{K} P_{j,k}}{K}. \tag{4.25}$$

Each Bayesian residual, $r_{j,k}$ is plotted against the mean of the fitted probabilities for each pairwise comparison, P_j. The plot will show the distribution of the Bayesian residuals for each pairwise comparison. Johnson and Albert (1999) state that residual distributions located far from a residual value of 0 are considered to be possible outliers. It is not clear how this distance is determined; however in an example presented by Johnson and Albert (1999), the residual distributions that are considered outliers have a median of either greater than 0.4 or less than -0.4.

4.6 MCMC results

MCMC is used to simulate from the posterior distributions of each of the parameters in $\boldsymbol{\theta}$ for three different choices of prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1). A sample of 10000 iterations was generated from the MCMC for each model. The Markov Chain for both the logit and probit models appears to reach the stationary distribution for all three priors within the first 10000 iterations. A further sample of 10000 iterations was then generated from the MCMC. A summary of the second sample is shown in this section.

4.6.1 Logit Model: Comparing Posterior Inferences with MLE

Individual Parameters

Table 4.5 shows the mean and the 95% posterior intervals of the model parameters, for the three prior distributions. The posterior means derived from all three priors are different to the maximum likelihood estimates of each parameter.

Consider the results using the Gamma(1,10) prior. Parameters with very small maximum likelihood estimates have larger posterior means. An example is the parameter Concern at level 2. The maximum likelihood estimate is 0.0013 but the posterior mean is 0.0216. However the 95% posterior interval is $(0.0008, 0.0638)$ which includes the maximum likelihood estimate. Parameters with larger maximum likelihood estimates tend to have smaller posterior means using a Gamma(1,10) prior. The parameter with the largest posterior mean is the same as the parameter with the largest maximum likelihood estimate: the attribute Activities at level 4. The maximum likelihood estimate of this parameter is 0.2420 and the posterior mean is 0.1927. The posterior interval is $(0.1236, 0.2556)$ which includes the maximum likelihood estimate.

Attribute	Level	Prior		
		Gamma(1,10)	Gamma(5,15)	Uniform(0,1)
Concern	2	0.0216(0.0008,0.0638)	0.0354 (0.0124,0.0681)	0.0200 (0.0006,0.0589)
	3	0.0175(0.0006,0.0558)	0.0349(0.0128,0.0644)	0.0177 (0.0005 ,0.0547)
	4	0.1049(0.0526,0.1580)	0.0633(0.0310,0.0984)	0.1008 (0.0440 ,0.1502))
	5	0.0344(0.0023,0.0816)	0.0548(0.0256,0.0882)	0.0384 (0.0031 ,0.0894)
Breath	2	0.0174(0.0006,0.0527)	0.0323(0.0115,0.0630)	0.0165(0.0006,0.0484)
	3	0.0182(0.0007,0.0560)	0.0348(0.0128,0.0641)	0.0189 (0.0007 ,0.0572)
	4	0.0903(0.0425,0.1347)	0.0598(0.0292,0.0934)	0.0920 (0.0445 ,0.1359)
	5	0.0177(0.0004,0.0584)	0.0366(0.0148,0.0693)	0.0185 (0.0006 ,0.0562)
Weather	2	0.0081(0.0002,0.028)	0.0216(0.0071,0.0409)	0.008(0.0002 ,0.029)
	3	0.0274(0.0012,0.0702)	0.0352(0.0138,0.0607)	0.0262 (0.0008 ,0.0676)
	4	0.0579(0.0087,0.1085)	0.0447(0.0190,0.0756)	0.0592 (0.0118 ,0.1084)
	5	0.0343 (0.0014,0.0938)	0.0445(0.0177,0.0798)	0.0353 (0.0018 ,0.0907)
Sleep	2	0.0149(0.0005,0.0496)	0.0277(0.0099,0.0522)	0.0147 (0.0005 ,0.0452)
	3	0.0330(0.0026,0.0771)	0.0336(0.0127,0.0599)	0.0337 (0.0023 ,0.0797)
	4	0.0296(0.0015,0.0765)	0.0352(0.0141,0.0649)	0.0315 (0.0018 ,0.0790)
	5	0.0287(0.0014,0.0722)	0.0354(0.0132,0.0645)	0.0280 (0.0012 ,0.0707)
Activities	2	0.0314 (0.0017,0.0820)	0.0436(0.0162,0.0782)	0.0303 (0.0013 ,0.0786)
	3	0.0391(0.0020,0.1000)	0.0625(0.0273,0.1048)	0.0421 (0.0025 ,0.0950)
	4	0.1927 (0.1236,0.2556)	0.1169(0.0670,0.1683)	0.1919 (0.1210 ,0.2579)
	5	0.0296(0.0011,0.0821)	0.0545(0.0224,0.0930)	0.0307 (0.0015 ,0.0806)
Scale Parameter		0.2366(0.2151,0.2626)	0.2264(0.2048,0.2488)	0.2386 (0.2167 ,0.2625)

Table 4.5: Mean and 95% posterior intervals for logit model by prior distribution

Health State Utilities

Using the 10000 sample vectors of parameters, $\boldsymbol{\theta}_1 0001, ..., \boldsymbol{\theta}_{20000}$, utilities are calculated for the 48 health states used in the sample survey. The utilities using the sample vector of parameters $\boldsymbol{\theta}_m$ are calculated by

$$g_m(\boldsymbol{x}_{jk}) = 1 - \boldsymbol{x}_{jk}^T \boldsymbol{\theta}_m, \quad j = 1, ...24, \quad k = 1, 2, \quad m = 10001, ...20000. \quad (4.26)$$

Figures 4.5, 4.6 and 4.7 present the mean and 95% posterior intervals of the 48 health states using the prior distributions Uniform(0,1), Gamma(1,10) and Gamma(5,15) respectively. The mean utilities for each of the 48 health states

calculated using the logit maximum likelihood estimates are also plotted. The utilities are plotted in decreasing order of the utility calculated using the maximum likelihood estimates. When a Uniform(0,1) or Gamma(1,10) prior is used, the posterior means of each health state are less than the utilities calculated using the maximum likelihood estimates. However the maximum likelihood estimates are included in the 95% posterior intervals. Most of the utilities calculated using the maximum likelihood estimates are not included in the 95% posterior intervals for the Gamma(5,15) prior.

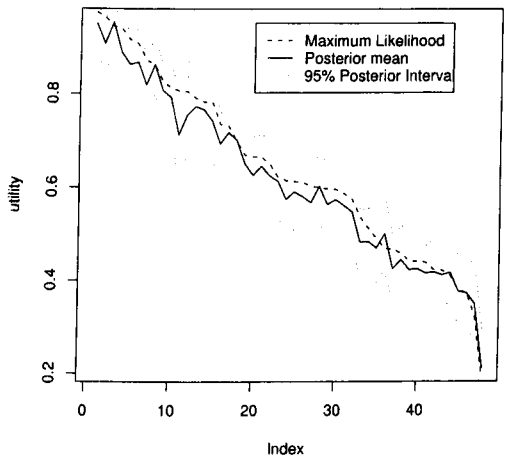


Figure 4.5: Mean utilities of 48 health states assuming a Uniform (0,1) prior

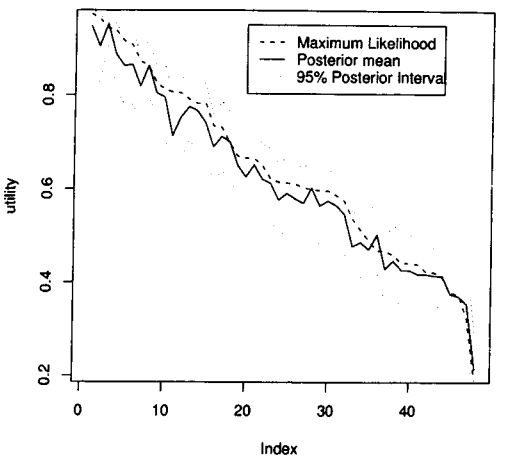


Figure 4.6: Mean utilities of 48 health states assuming a Gamma(1,10) prior

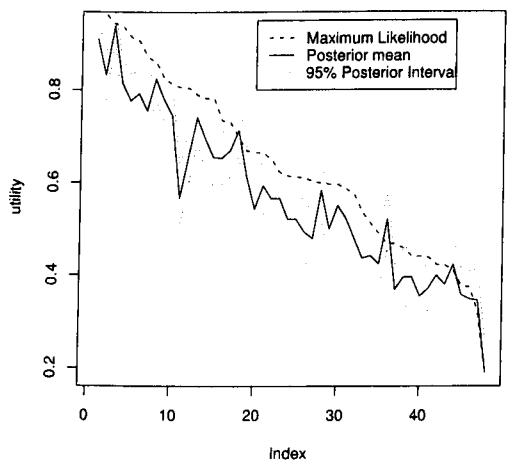


Figure 4.7: Mean utilities of 48 health states assuming a Gamma(5,15) prior

4.6.2 Logit Model: Comparing Priors

Individual Parameters

Table 4.5 also shows the posterior mean and 95% posterior intervals of the parameters for each prior distribution. The posterior mean and 95% probability intervals for most parameters are similar for the both Uniform(0,1) and Gamma(1,10) prior. The posterior mean and 95% posterior intervals are significantly changed using the Gamma(5,15) prior distribution. When the posterior means in the Gamma(1,10) and Uniform(0,1) models are small the posterior means in the Gamma(5,15) model are increased. When the posterior means in the Gamma(1,10) and Uniform(0,1) models are larger the posterior means in the Gamma(5,15) are decreased. These observations are also shown by the posterior distributions of each parameter which are presented in Appendix B. Posterior distributions are plotted for each parameter given the three prior distributions. Using a Uniform(0,1) and Gamma(1,10) prior distribution produces similar posterior distributions. Most of these posterior distributions have a large probability of parameter values being close to zero. Using a Gamma(5,15) prior distribution produces a posterior which does not allow a parameter value close to zero. Parameters that represent a change from level 3 to level 4, such as parameter 3 and parameter 7, have posterior distributions further from zero. The posterior distributions of parameter 3 and parameter 7 have a smaller mode for a Gamma(5,15) prior than when either a Uniform(0,1) or Gamma(1,10) prior is used.

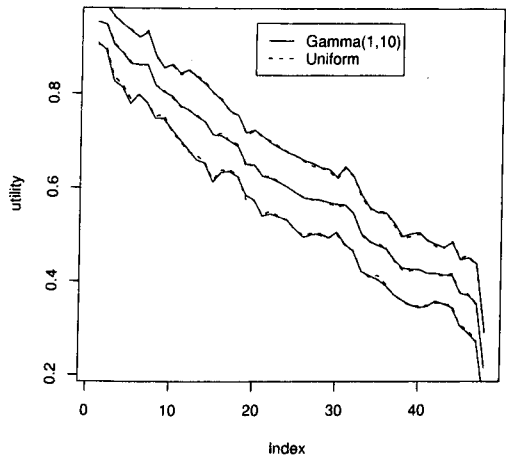


Figure 4.8: Posterior mean and 95% posterior intervals for 48 health states using Gamma(1,10) and Uniform(0,1) priors

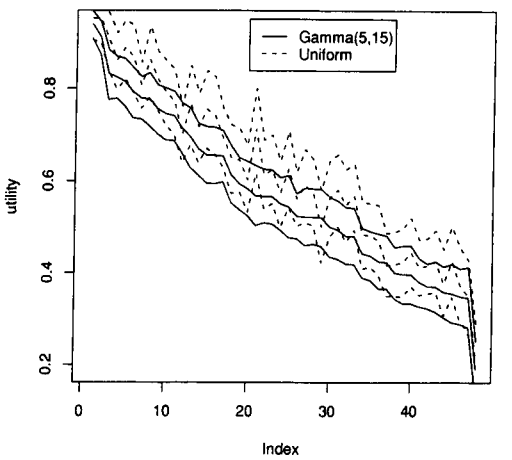


Figure 4.9: Posterior mean and 95% posterior intervals for 48 health states using Gamma(5,15) and Uniform(1,10) priors

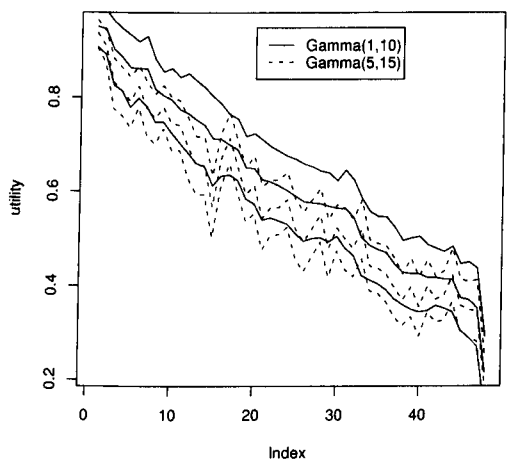


Figure 4.10: Posterior mean and 95% posterior intervals for 48 health states using Gamma(1,10) and Gamma(5,15) priors

Health State Utilities

Figure 4.8 presents the posterior means and 95% posterior intervals for each of the 48 health states used in the sample survey, given a Uniform(0,1) prior and a Gamma(1,10) prior. The utilities are plotted in decreasing order of the posterior mean from the uniform(0,1) prior. Figures 4.9 and 4.10 compare the posterior means and 95% posterior intervals from the Gamma(5,15) prior, with those from the uniform(0,1) and Gamma(1,10) priors. The mean utilities using a Uniform(0,1) and a Gamma(1,10) prior are very similar. The mean utilities using a Gamma(5,15) prior are usually smaller than those from the uniform(0,1) and Gamma(1,10) priors.

Figure 4.11 presents the posterior distributions of the worst health state defined by the AQL-5D classification system, for each of the three priors. The maximum likelihood estimate of the health state is also presented. Using a Uniform(0,1) prior the posterior distribution has mean 0.1450 and 95% posterior interval $(0.0963, 0.1926)$. Using a Gamma(1,10) prior, the posterior distribution has mean 0.1504 and 95% posterior interval $(0.1019, 0.1991)$. Using a Gamma(5,15) prior, the posterior distribution has mean 0.0920 and 95% posterior interval $(0.0458, 0.1379)$. The posterior distributions are very similar when a Gamma(1,10) or Uniform(0,1) prior is used. When a Gamma(5,15) prior is used, the mode is decreased and the distribution is closer to zero.

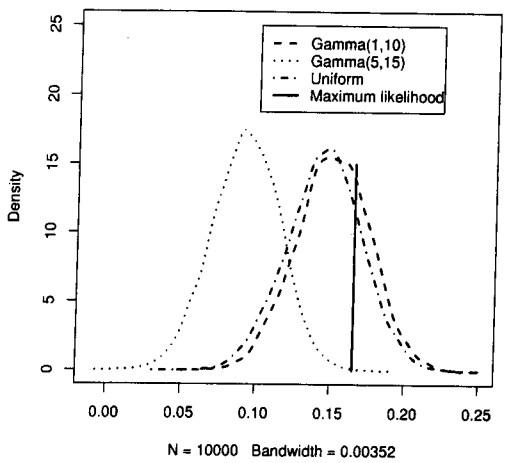


Figure 4.11: Posterior Distributions of worst health state

Residuals

Figures 4.12, 4.13 and 4.14, on page 98, present the Bayesian residuals using the Uniform(0,1) prior, Gamma(1,10) prior and Gamma(5,15) prior respectively. Using my observation of the example presented in Johnson and Albert (1999) where the outliers have a median or either greater than 0.4 or less than -0.4, I have concluded that there are no outliers in the residual plots for the three prior distributions.

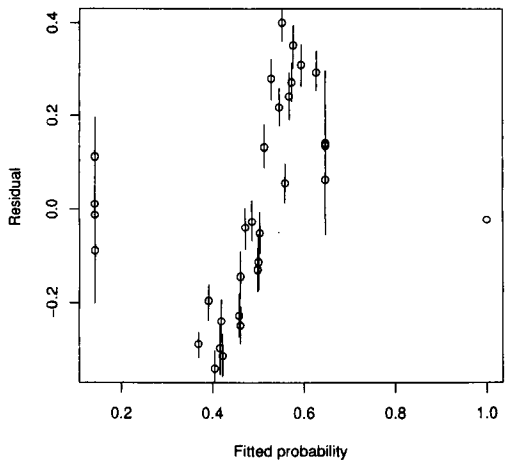


Figure 4.12: Residuals Uniform (0,1) prior

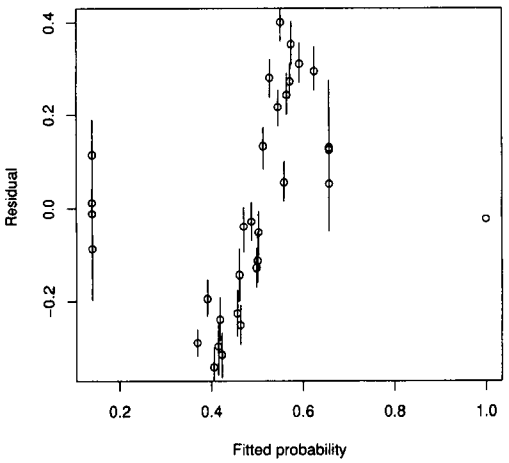


Figure 4.13: Residuals Gamma(1,10) prior

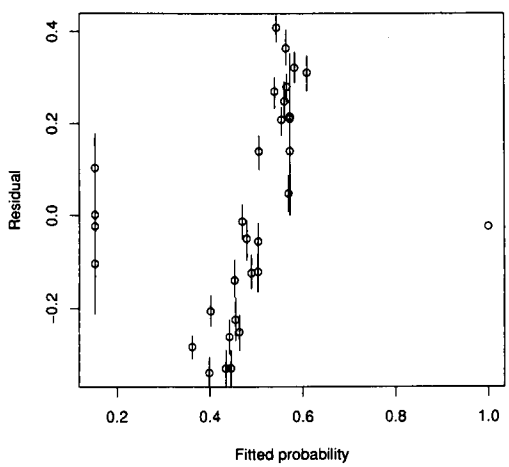


Figure 4.14: Residuals Gamma(5,15) prior

4.6.3 Comparing Logit and Probit Likelihoods

Figures 4.15, 4.16 and 4.17, on page 100 compare the mean utilities and 95% posterior intervals calculated from the logit and probit models for the three prior

distributions: Uniform(0,1), Gamma(1,10) and Gamma(5,15). The models look most similar when a Gamma(5,15) prior is used.

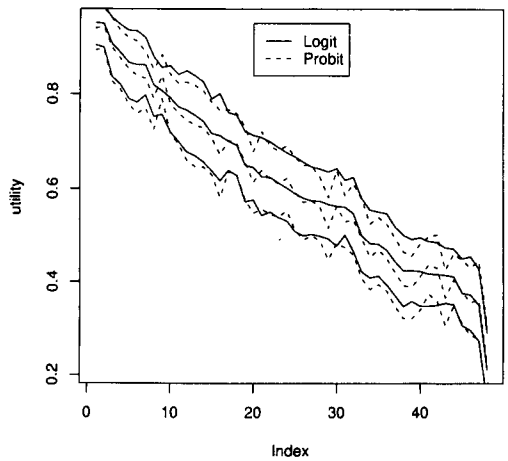


Figure 4.15: Mean and 95% posterior intervals for 48 health states assuming a Uniform(0,1) prior

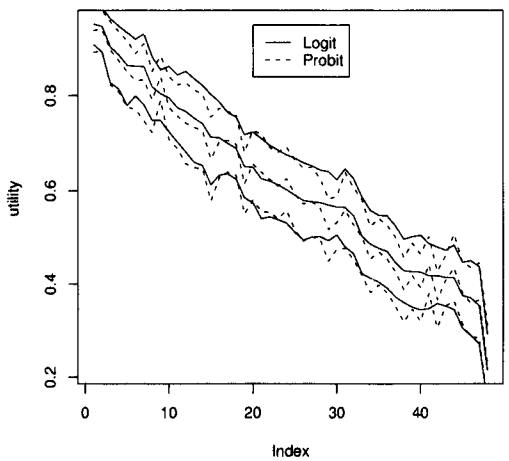


Figure 4.16: Mean and 95% posterior intervals for 48 health states assuming a Gamma(1,10) prior

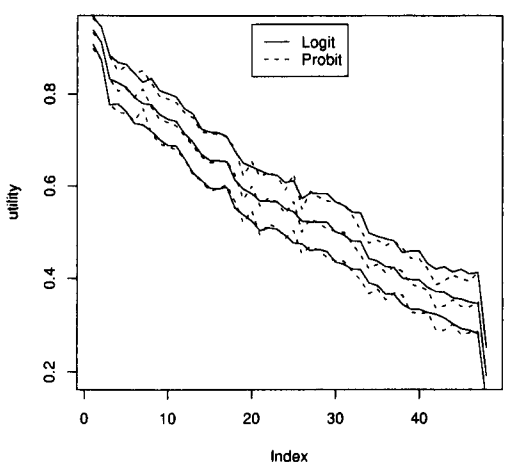


Figure 4.17: Mean and 95% posterior intervals for 48 health states assuming a Gamma(5,15) prior

4.7 Comparison with TTO model

Yang et al. (2006) presents an analysis of TTO data collected for health states defined using the AQL-5D classification system. Each respondent in the study valued 8 health states including the worst possible health defined by the AQL-5D classification system. Several models were considered of the form

$$y_{ij} = g(\beta^T \boldsymbol{x}_{ij} + \theta^T \boldsymbol{r}_{ij} + \delta^T \boldsymbol{z}_j) + \varepsilon_{ij}, \tag{4.27}$$

where y_{ij} is the TTO score for health state i valued by respondent j, $\boldsymbol{x}$ is a vector of binary dummy variables defining the health state, $\boldsymbol{z}$ is a vector of personal characteristics such as sex, age and asthma condition, and $\boldsymbol{r}$ is a vector of interactions between attributes.Yang et al. (2006) determined the most appropriate model to be represented by the parameter estimates in table 4.6. There are no demographic or interaction variables considered in the model. Each parameter estimate represents the incremental decrease in utility when an attribute increases by one level. This is the same as in the discrete choice models. Some attribute levels ere combined due to some inconsistent parameter estimates in earlier models. Comparison with the probit and logit maximum likelihood results shows a smaller range between the largest and smallest estimates and unlike the discrete choice models there are no parameter estimates of zero. The largest decrease in mean utility is a change from level 3 to level 4 for the attribute activities. This is the same as the logit and probit models. However the decrease is larger in the logit and probit models. The utility of the worst possible health state defined by the model in Yang et al. (2006) is -0.0566. Assuming that the health state death has a utility of 0 this implies the worst health state is worst than death. However, it is not clear in Yang et al. (2006) how the utility of death is defined.

Attribute	Concern	Breath	Weather	Sleep	Activities
Level 2	0.027	0.030	0.0103	0.041	0.011
Level 3	0.046	-	0.028	0.056	0.065
Level 4	0.064	-	0.058	0.073	0.177
Level 5	0.064	0.106	0.113	0.090	-

Table 4.6: TTO model parameter estimates

4.8 Conclusion

When a Gamma(1,10) prior distribution is used the mean utilities and posterior distributions of the parameters are similar to those when a uniform(0,1) prior is used. When a Gamma(5,15) prior distribution is used, the results change. If the prior distribution does not favour larger values the posterior distribution is robust to the prior. Other prior distributions could be investigated. For most parameters where the value would be expected to be quite small, a prior that favours smaller values would be appropriate. However, some parameters might be expected to reduce utility more than other parameters and a prior that favours larger values might be more appropriate.

Further MCMC analyses were performed where different priors were considered for the scale parameter, σ, while keeping the priors on the parameter $\boldsymbol{\theta}$ constant. In all analyses the posterior distribution of the scale parameter and the 95% posterior interval of the utilities did not change when changing the prior on the scale parameter.

The utilities calculated using the maximum likelihood estimates tend to be greater than the posterior means of the utilities for all three priors. This suggests a skewed distribution. The posterior intervals present an estimate of the uncertainty in the distribution of utilities. The posterior intervals of the 48 health states, assuming either a Gamma(1,10) or a Uniform(0,1) prior, are between approximately 0.1

and 0.2 for both the logit and probit models. This is presented in figures 4.15 and 4.16. Considering the utilities are between 0 and 1, this is fairly small. The posterior interval appears to be smaller when using a Gamma(5,15) prior, as presented in figure 4.17. The uncertainty in the posterior distribution of the utilities is the same for both logit and probit models.

When preference is given to smaller values in the prior distribution the results do not change from when all parameter values are equally likely. Results change when preference is given to larger parameter values in the prior and smaller values are less likely.

The use of Bayesian inference allows the assessment of uncertainty of each parameter in the model. For a given health state the uncertainty in the utility can be derived by combining the posterior distributions of each parameter relevant to the health state. When investigating the cost-effectiveness analysis of a treatment, a distribution of possible changes in QALY can be derived. As described in chapter 1, NICE requires the reporting of the probability that the treatment is cost-effective at the threshold of $£20,000 - £30,000$ per QALY and the probability that the treatment is not effective. The distribution of possible changes in QALY and the uncertainty in cost can be combined to produce a distribution from which the required probabilities can be derived.

Chapter 5

Comparing Models using Bayes Factors

5.1 Introduction

In chapter 4, we fitted several Bayesian models to the AQL-5D discrete choice data. The results showed some differences between the models but did not determine which model was the most suitable for the data. Here we consider the use of Bayes factors for model comparison.

The logit and probit models are compared for each of the three prior distributions used in chapter 4. To understand and interpret the values of these Bayes factors, the logit and probit models are both compared to a benchmark model. In the benchmark model the attributes are assumed to be irrelevant to the choices made in each comparison and the preferred health state in each pair is chosen randomly.

We also investigate a hypothesis that respondents do not consider all five attributes when choosing a preferred health state. Five new models are fitted where each attribute has been removed from the data set. Bayes factors are calculated

comparing each of these models with the full model containing all five attributes. The value of each Bayes factor shows the importance of each attribute on the full model.

5.2 Definition of Bayes factors

Kass and Raftery (1995) review the use of Bayes factors, which can be used as a method of model comparison. Kass and Raftery (1995) reference two articles, Jeffreys (1935) and Jeffreys (1961) which developed the Bayesian approach to hypothesis testing. Statistical models are introduced to represent the probability of the data for each of the two hypotheses and Bayes theorem is then used to calculate the posterior probability that one of the hypotheses is correct. Kass and Raftery (1995) describe Bayes factors as a way of evaluating evidence in favour of a null hypothesis.

Let M_1 and M_2 be the two models being compared. Define $\boldsymbol{\theta}_1$ to be the set of parameters for model M_1 and $\boldsymbol{\theta}_2$ to be the set of parameters for model M_2. The marginal likelihood of the data $\boldsymbol{y}$ given the model M_k, $P(\boldsymbol{y}|M_k)$, $k = 1, 2$ is calculated using

$$P(\boldsymbol{y}|M_k) = \int_{\Theta_k} P(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k) P(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k, \tag{5.1}$$

where $P(\boldsymbol{\theta}_k|M_k)$ is the prior density of the set of parameters, $\boldsymbol{\theta}_k$, for model M_k, $P(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)$ is the likelihood of the data $\boldsymbol{y}$ given the vector of parameters $\boldsymbol{\theta}_k$ and Θ_k is the set of all possible vectors $\boldsymbol{\theta}_k$. Define $P(M_1)$ and $P(M_2)$ to be the prior probabilities of model M_1 and M_2, and define $P(M_1|\boldsymbol{y})$ and $P(M_2|\boldsymbol{y})$ to be the posterior probabilities given data $\boldsymbol{y}$. The ratio of the posterior probabilities

for model M_1 and model M_2 is

$$\frac{P(M_2|\boldsymbol{y})}{P(M_1|\boldsymbol{y})} = \frac{P(M_2)}{P(M_1)} \quad \times \quad B_{21}. \tag{5.2}$$

The Bayes factor, B_{21}, is the ratio of the marginal likelihood for model M_2 to the marginal likelihood for model M_1, and is given by

$$B_{21} = \frac{P(\boldsymbol{y}|M_2)}{P(\boldsymbol{y}|M_1)}. \tag{5.3}$$

The Bayes factor, B_{21}, is a value which represents the evidence provided by the data in favour of model M_2 as opposed to model M_1. If the Bayes factor is greater than 1, then the data favour model M_2, and if the Bayes factor is less than 1 the data favour model M_1. The value of the Bayes factor determines the strength of evidence in favour of model M_2. Jeffreys (1961) suggested a set of categories to interpret the value of a Bayes factor, which are also reviewed by Gelman et al. (2004), Robert (2001) and Kass and Raftery (1995). They are summarised as follows. If the Bayes factor, B_{21}, is above 100, there is decisive evidence against model M_1. If B_{21} is between 10 and 100 there is strong evidence and if B_{21} is between 3 and 10 there is substantial evidence against model M_1. Kass and Raftery (1995) suggest that interpretation of Bayes factors may depend on the context and use Evett (1991) as an example, where it is argued that for forensic evidence alone to be decisive in a criminal trial the posterior odds, in favour of guilt against innocence, would be required to be at least 1000.

5.3 Calculation of Bayes factors

Kass and Raftery (1995) and Robert (2001) discuss several methods to calculate Bayes factors. In this chapter, importance sampling is used to calculate Bayes factors. Robert (2001) reviewed importance sampling and considers the method to be well adapted to the computation of predictive distributions such as the

marginal likelihood in equation (5.1). We now describe the method, following the description in Robert (2001). A probability $p(y)$ can be calculated using Monte Carlo integration,

$$p(y) = \int p(y|\theta)p(\theta)d\theta, \tag{5.4}$$

where $p(\theta)$ is the density of a random variable θ and $p(y|\theta)$ is the conditional probability of y given θ. If a random sample of parameters $\theta_1, ..., \theta_N$ is generated from $p(\theta)$, $p(y)$ can be estimated using

$$\hat{p}(y) = \frac{1}{N}\sum_{n=1}^{N} p(y|\theta_n). \tag{5.5}$$

It can be observed that equation (5.4) is the expectation w.r.t θ of a function $p(y|\theta)$ where θ has density $p(\theta)$.

Equation (5.4) can be estimated using a sample $\theta_1, ..., \theta_N$ even if the sample has not been generated from $p(\theta)$. In importance sampling, a sample of parameters $\theta_1, ..., \theta_N$ are generated from a density $h(\theta)$, called the importance density, and the probability $p(y)$ in equation (5.4) is estimated by

$$\hat{p}(y) = \frac{1}{N}\sum_{n=1}^{N}\left(p(y|\theta_n)\frac{p(\theta_n)}{h(\theta_n)}\right). \tag{5.6}$$

The sample $\theta_1, ..., \theta_N$ is a weighted sample where each parameter θ_i has weighting $\frac{p(\theta_n)}{h(\theta_n)}$. It can be observed that equation (5.6) is a valid estimate of equation (5.4) by writing

$$p(x) = \int \left[p(y|\theta)\frac{p(\theta)}{h(\theta)}\right] h(\theta)d\theta, \tag{5.7}$$

i.e. $p(y)$ can be expressed as the expectation of $p(y|\theta)\frac{p(\theta)}{h(\theta)}$, where θ has density $h(\theta)$.

Importance sampling can be used to estimate the likelihood $P(\boldsymbol{y}|M_k)$ in equation

(5.1). Geweke (1989) states the following assumptions that are required.

1 The product of the prior density, $P(\boldsymbol{\theta}_k|M_k)$, and the likelihood, $P(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)$, is proportional to a proper probability density function defined on $\boldsymbol{\Theta}$

2 $\{\boldsymbol{\theta}_i\}_{i=1}^{\infty}$ is a sequence of i.i.d random vectors, the common distribution having a probability density function $h(\boldsymbol{\theta}_k)$

3 The support of $h(\boldsymbol{\theta}_k)$ includes $\boldsymbol{\Theta}$

4 $P(\boldsymbol{y}|M_k)$ exists and is finite

If a sample of parameter vectors $\boldsymbol{\theta}_k^{(1)}, ..., \boldsymbol{\theta}_k^{(N)}$ is generated from the importance distribution $h(\boldsymbol{\theta}_k)$, equation (5.1) is estimated by

$$\hat{P}(\boldsymbol{y}|M_k) = \sum_{n=1}^{N} \left(P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k) \frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{h(\boldsymbol{\theta}_k^{(n)})} \right). \tag{5.8}$$

The choice of distribution for $h(\boldsymbol{\theta}_k)$ determines the equation for the estimator $\hat{P}(\boldsymbol{y}|M_k)$. If $h(\boldsymbol{\theta}_k)$ is equal to the prior distribution , $h(\boldsymbol{\theta}_k) = P(\boldsymbol{\theta}_k^{(n)}|M_k)$, then

$$\hat{P}(\boldsymbol{x}|M_k) = \frac{1}{N} \sum_{n} P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k). \tag{5.9}$$

Robert (2001) states that the estimator in equation (5.9) is often inefficient if the data is informative because most simulated values of $\boldsymbol{\theta}_k^{(n)}$ will be outside the modal region of the likelihood. A more efficient method is to let $h(\boldsymbol{\theta}_k)$ be equal to the posterior distribution, $h(\boldsymbol{\theta}_k) = P(\boldsymbol{\theta}_k|\boldsymbol{y}, M_k)$. Then

$$\hat{P}(\boldsymbol{y}|M_k) = \sum_{n=1}^{N} \left(P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k) \frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{y}, M_k)} \right). \tag{5.10}$$

However, since

$$P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{y}, M_K) = \frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)}{p(\boldsymbol{y}|M_k)}, \tag{5.11}$$

we cannot evaluate $P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{y}, M_k)$, without already knowing the value of $P(\boldsymbol{y}|M_k)$, the probability we are trying to estimate. Using the fact that the integral of the prior density over the parameter space is $\int P(\boldsymbol{\theta}_k|M_k)d\theta_k = 1$, to calculate the marginal likelihood in equation (5.1), we first re-write it as

$$P(\boldsymbol{y}|M_k) = \frac{\int_\Theta P(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)P(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}{\int_\Theta P(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}. \tag{5.12}$$

We now consider importance sampling to estimate both numerator and denominator simultaneously. If a sample of vectors $\boldsymbol{\theta}_k^{(1)}, ..., \boldsymbol{\theta}_k^{(N)}$ is generated from the importance distribution $h(\boldsymbol{\theta}_k)$, equation (5.12) is estimated by

$$\hat{P}(\boldsymbol{y}|M_k) = \frac{\sum_{n=1}^{N}\left(P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)\frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{h(\boldsymbol{\theta}_k^{(n)})}\right)}{\sum_{n=1}^{N}\left(\frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{h(\boldsymbol{\theta}_k^{(n)})}\right)}. \tag{5.13}$$

When the importance distribution used to generate each sample is the posterior distribution, $P(\boldsymbol{\theta}_k^{(n)}|M_k)$, we have

$$\hat{P}(\boldsymbol{x}|M_k) = \frac{\sum_{n=1}^{N}\left(P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)\frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{y}, M_k)}\right)}{\sum_{n=1}^{N}\left(\frac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{x}, M_k)}\right)}. \tag{5.14}$$

As $P(\boldsymbol{\theta}_k^{(n)}|\boldsymbol{y}, M_k) = \dfrac{P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{y}|M_k)}$, we now have

$$\begin{aligned}
\hat{P}(\boldsymbol{y}|M_k) &= \frac{\sum_{n=1}^{N}\left(P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)\dfrac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)P(\boldsymbol{\theta}_k^{(n)}|M_k)}\right)}{\sum_{n=1}^{N}\left(\dfrac{P(\boldsymbol{\theta}_k^{(n)}|M_k)}{P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)P(\boldsymbol{\theta}_k^{(n)}|M_k)}\right)} \\
&= \frac{\sum_{n=1}^{N} 1}{\sum_{n=1}^{N}\left(\dfrac{1}{P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)}\right)} \\
&= \frac{N}{\sum_{n=1}^{N}\left(\dfrac{1}{P(\boldsymbol{y}|\boldsymbol{\theta}_k^{(n)}, M_k)}\right)}.
\end{aligned} \tag{5.15}$$

Suppose N vectors of parameters are sampled from their respective posterior distributions for both model M_1 and model M_2. Using equations (5.3) and (5.15), the Bayes factor comparing model M_1 and model M_2 is calculated using

$$B_{21} = \frac{\left(\dfrac{N}{\sum_{n=1}^{N}\left(\dfrac{1}{P(\boldsymbol{y}|\boldsymbol{\theta}_2^{(n)}, M_2)}\right)}\right)}{\left(\dfrac{N}{\sum_{n=1}^{N}\left(\dfrac{1}{P(\boldsymbol{y}|\boldsymbol{\theta}_1^{(n)}, M_1)}\right)}\right)}. \tag{5.16}$$

We can obtain the parameter samples using MCMC, as described in section 4.5.2.

5.3.1 BENCHMARK MODEL

Bayes factors can be used to compare the probit and logit models. If M_1 is the logit model and M_2 is the probit model, the value of the Bayes factor in equation (5.16) represents the level of evidence in favour of the probit model. To understand the value of the Bayes factor and appreciate the strength of the evidence it represents, it is useful to define a benchmark model, which we consider to be the simplest possible model for the data. Bayes factors can then be calculated comparing the probit and logit models with the benchmark model. The improvement of the probit model over the logit model, represented by a Bayes factor value can be compared with the improvement of the logit model and the probit model over the benchmark model.

The benchmark model assumes that the level of each attribute is irrelevant to the choice made in each health state comparison. In pairwise choice data, an individual chooses the preferred health state from a set $B = \{\boldsymbol{x}_{j1}, \boldsymbol{x}_{j2}\}$. Each individual is making a decision based on the preferred combination of attribute levels that define each health state. If an individual has no preference between each level of the attributes then the two health states in the comparison have an equal chance of being chosen. The choice can then be considered random. Define $P(\boldsymbol{x}_{j1})$ to be the probability that an individual prefers health state $\boldsymbol{x}_{j1}$ to health state $\boldsymbol{x}_{j2}$. If model M_1 assumes that choices are random then $P_B(\boldsymbol{x}_{j1}) = 0.5$ for all pairwise comparisons and the marginal likelihood is calculated by

$$P(\boldsymbol{x}|M_1) = \prod_{j=1}^{J}(0.5)^{N_j}. \tag{5.17}$$

where N_j is the number of individuals comparing health state $\boldsymbol{x}_{j1}$ and health state $\boldsymbol{x}_{j2}$. If model M_1 is the benchmark model, the Bayes factor, B_{21}, is

$$B_{21} = \frac{\left(\frac{N}{\sum_{n=1}^{N} \left(\frac{1}{P(\boldsymbol{y}|\boldsymbol{\theta}_2^{(n)}, M_2)} \right)} \right)}{\prod_{j=1}^{J} (0.5)^{N_j}}. \tag{5.18}$$

5.4 Initial Results

MCMC simulations are run for the logit and probit models using the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1). Section 5.4.1 presents Bayes factors comparing the logit model with the benchmark model, section 5.4.2 presents Bayes factors comparing the probit model with the benchmark model and section 5.4.3 presents Bayes factors comparing the logit and probit models. When using notation, let M_L represent the logit model, M_P represent the probit model and M_B represent the benchmark model. Define $P(N|M_L)$, $P(N|M_P)$ and $P(N|M_B)$ to be the probability of the data, N, given each of the respective models. A sample of 210000 iterations are generated using MCMC for each model. The first 10000 are discarded as burn-in and the remaining 200000 are summarised in this section.

5.4.1 Comparing the logit model with the benchmark model

This section presents the Bayes factors supporting evidence in favour of the logit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1). Bayes Factors are calculated using equation (5.18) where model M_2 is the logit model with each of the three prior distributions. The Bayes factors are presented in table 5.1.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	1.4867×10^{109}	1.2000×10^{109}	8.6429×10^{108}	8.4935×10^{108}
Gamma(5,15)	1.5124×10^{103}	5.4430×10^{102}	7.8604×10^{102}	7.8810×10^{102}
Uniform(0,1)	5.7160×10^{108}	5.9451×10^{108}	5.2381×10^{108}	4.7956×10^{108}

Table 5.1: $P(N|M_L)/P(N|M_B)$

Using the guidelines reviewed in section 5.2 the Bayes factors show decisive evidence in favour of the logit model. Considering the size of the Bayes factors for all three priors the three logit models are significantly better than the benchmark model. The model using the Gamma(5,15) is the least favoured out of the three prior distributions but this is only by a small amount relative to the size of the Bayes factors. The Bayes factors remain fairly constant over iterations.

5.4.2 Comparing the probit model with the benchmark model

This section presents the Bayes factors supporting evidence in favour of the probit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1). Bayes factors are calculated using equation (5.18) where model M_2 is the probit model with each of the three prior distributions. The Bayes factors are presented in table 5.2.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	1.1550×10^{117}	9.3886×10^{116}	1.1806×10^{117}	9.8616×10^{116}
Gamma(5,15)	1.4828×10^{113}	8.9092×10^{112}	1.0095×10^{112}	1.1757×10^{112}
Uniform(0,1)	6.1936×10^{116}	9.3633×10^{116}	1.1514×10^{117}	1.0273×10^{117}

Table 5.2: $P(N|M_P)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the probit model. As with the Bayes factors in section 5.4.1 for the logit models, the size of the Bayes factors for all three priors suggests the three probit models are significantly better than the benchmark model. Again, as with the logit model, the probit model assuming a Gamma(5,15) is the least favoured out of the three prior distributions but this is only by a small amount relative to the size of the Bayes factors. The Bayes factors remain fairly constant over iterations.

5.4.3 Comparing the logit model with the probit model

This section presents Bayes factors to evaluate evidence in favour of the probit Model over the logit model. The Bayes factors using three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1), are presented in table 5.3. Bayes factors are calculated using equation (5.16) where M_1 is the logit model and model M_2 is the probit model.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	7.7688×10^7	7.8241×10^7	1.3660×10^8	1.1611×10^8
Gamma(5,15)	9.8040×10^9	1.6368×10^{10}	1.2843×10^9	1.4918×10^9
Uniform(0,1)	1.0836×10^8	1.5749×10^8	2.1982×10^8	2.1422×10^8

Table 5.3: $P(N|M_P)/P(N|M_B)$

The Bayes factors presented in table 5.3 show the probit model is a better fit for the data than the logit model. Using the rules defined in Section 5.2 there is decisive evidence in favour of the probit model. The Bayes factors are a lot larger than the threshold value for decisive evidence that it suggests that the probit model is a much better fitting model for the data than the logit model.

5.4.4 Discussion

Bayes factors showing evidence in favour of both the logit and probit models when compared with the benchmark model are very high which suggests that the respondents answering the DCE questions are not selected health states randomly and therefore either a logit model or a probit model is a more appropriate representation of the AQL-5D health state utilities. The Bayes factors in favour of the probit model against the logit model are fairly large. Relative to the values of the Bayes factors comparing the probit and logit models with the benchmark model, they are of a reasonable size. However, it is still surprising that there is a difference between the two models. The assumed error distributions, the normal distribution and the type 1 extreme value distribution are very similar. The results in chapter 4 also show no difference between the two models. Figures 4.14, 4.15 and 4.16 show the mean and posterior intervals for the probit and logit models using each of the three prior distributions. These graphs show little difference between the models.

To understand the differences between the models, values of the likelihoods were calculated for the logit and probit models using both the maximum likelihood estimates and samples of parameters generated using MCMC. The logit and probit likelihood equations defined in section 4.3 are then separated into two subsets. The likelihoods are now written in the form

$$P(\boldsymbol{N}|\boldsymbol{\theta},\sigma^2,M) = P_1(\boldsymbol{N}|\boldsymbol{\theta},\sigma^2,M) + P_2(\boldsymbol{N}|\boldsymbol{\theta},\sigma^2,M), \qquad (5.19)$$

where $P_1(\boldsymbol{N}|\boldsymbol{\theta},\sigma^2,M)$ is the likelihood for pairwise comparisons that do not include the health state death and $P_2(\boldsymbol{N}|\boldsymbol{\theta},\sigma^2,M)$ is the likelihood for pairwise comparisons that include the health state death. Values of P_1 and P_2 are calculated for the logit and probit models using maximum likelihood estimates and

samples of parameter values generated using MCMC. When comparing the values of P_1 for the logit and probit models there was little difference. However the value of the probit model for P_2 was significantly larger than that for the logit. After further investigation it was found that this larger difference was due to data collected for the comparison between death and the health state $(5, 2, 1, 1, 1)$.

For comparisons including the health state death, the probit model is still symmetrical but the logit model has a distribution which is has a positive skew and the probability of preferring a health state $\boldsymbol{x}_{i1}$ to death is given by

$$P_B(\boldsymbol{x}_{i1}) = 1 - \exp\left(-\exp\left(\frac{g(\boldsymbol{x}_{i1}) - 0.5722\sigma}{\sigma}\right)\right). \tag{5.20}$$

For most of the health states compared with death this skewness does not effect the likelihood values. However, the unusual choice of the health state $(5, 2, 1, 1, 1)$ to be used as a comparison with death may have an effect on the difference between the likelihoods. To understand the reason for this difference between the probit and logit likelihoods it is useful to draw the probability distributions for the health state $(5, 2, 1, 1, 1)$ using both the logit and probit models.

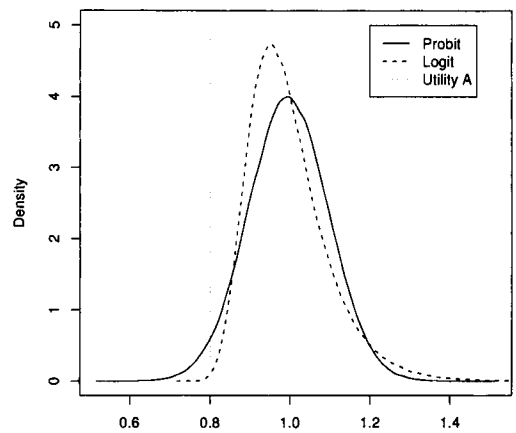


Figure 5.1: Probability distributions of health state (5,2,1,1,1)

Figure 5.1 presents the distributions of the utility of health state $(5, 2, 1, 1, 1)$ using the logit and probit models. In the AQL-5D discrete choice experiment questionnaire one comparison is between death and the health state $(5, 2, 1, 1, 1)$ and out of 42 respondents evaluating this comparison one prefers death. The probability an individual prefers death to health state $(5, 2, 1, 1, 1)$ is equal to the probability that the utility of health state $(5, 2, 1, 1, 1)$ is less than the utility of death. It can be observed from figure 5.1 that the logit model has a positive skew and the probit model is symmetrical. The utility that the individual who preferred death has for the other health state in the comparison, $(5, 2, 1, 1, 1)$, will be in the tail of the distributions. An example of such a utility is labelled as Utility A in the graph in figure 5.1. The density to the left of this line is significantly larger for the probit model than the logit model, which accounts for the difference in likelihood between the logit and probit and therefore the unexpectedly high Bayes factor.

To investigate the Bayes factor further, probabilities for each of the compar-

isons in the data set were calculated using a sample of parameter values. The probability of a respondent preferring health state $(5, 2, 1, 1, 1)$ to death is significantly larger when using the probit model compared to the logit model. For all other comparisons the probability is approximately equal. This difference in probability for one comparison between the logit and probit models explains the large Bayes factor in favour of the probit model. To investigate this further, the data set could be changed to either exclude the comparison between death and the health state $(5, 2, 1, 1, 1)$ or change the number of respondents preferring death to 0. The Bayes factors can then be calculated again and compared to the initial values. An alternative is to assume a negative error for the utility equation so utility, U_{i1}, of health state $\boldsymbol{x}_{i1}$ is then defined as

$$U_{ij} = g(\boldsymbol{x}_{ij}) - \varepsilon_{i1}. \tag{5.21}$$

This would not change the likelihood in the probit model. The logit likelihood will change for comparisons including death but will remain unchanged for other comparisons. The probability distribution for the comparisons involving death for the extreme value error would be skewed in the opposite direction which could affect the value of the Bayes factors. An investigation into the value of the Bayes factors when changes are made to the data set and negative errors are assumed is presented in Section 5.5.

5.5 Re-analysis of AQL-5D data

This section presents Bayes factor results when some changes are made to the probit and logit models. Bayes factors are again calculated comparing the logit and probit models with the benchmark model and comparing the probit with the logit model for each of the three prior distribution: Gamma(1,10), Gamma(5,15) and Uniform(0,1). For each new set of models samples of 210000 iterations are

generated using MCMC. The first 10000 iterations are discarded as burn-in and the remaining 200000 are used to calculate Bayes factors. Section 5.5.1 presents Bayes factors using a data set with the comparisons between death and health state $(5, 2, 1, 1, 1)$ removed, section 5.5.2 presents Bayes factors when the number of respondents choosing death in the comparison between health state $(5, 2, 1, 1, 1)$ is changed from 1 to 0 and section 5.5.3 presents Bayes factors when a negative error is assumed.

5.5.1 Bayes factors when one death comparison removed from data set

This section presents the Bayes factor results when the comparison between death and health state $(5, 2, 1, 1, 1)$ is removed from the initial AQL-5D discrete choice data set.

Comparing the logit model with the benchmark model

Bayes factors supporting evidence in favour of the logit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1) are presented in table 5.4.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	1.6442×10^{108}	1.4832×10^{108}	2.0279×10^{108}	2.6327×10^{108}
Gamma(5,15)	1.6643×10^{103}	2.5608×10^{103}	3.3599×10^{103}	4.3028×10^{103}
Uniform(0,1)	1.8745×10^{109}	1.1146×10^{109}	1.0764×10^{109}	8.3417×10^{108}

Table 5.4: $P(N|M_L)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the logit model over the random model. The Bayes factor values are very similar to those calculated using the initial data shown in table 5.1.

Comparing the probit model with the benchmark model

Bayes factors supporting evidence in favour of the probit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1) are presented in table 5.5.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	1.3787×10^{109}	1.3002×10^{109}	1.7837×10^{108}	4.5778×10^{107}
Gamma(5,15)	3.8093×10^{103}	6.0118×10^{103}	8.6671×10^{103}	9.1358×10^{103}
Uniform(0,1)	6.2877×10^{108}	7.1776×10^{108}	4.5398×10^{108}	2.5119×10^{108}

Table 5.5: $P(N|M_P)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the probit model over the benchmark model. The Bayes factor values are very different to those calculated using the initial data shown in table 5.2. Using a Gamma(1,10) prior, the values have reduced by a factor of 10^8, the model with a Gamma(5,15) prior is reduced by 10^{10} and the model with a Uniform(0,1) prior is reduced by a factor of 10^8

Comparing the probit model with the logit model

Bayes factors to evaluate evidence in favour of the probit model over the logit model using three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1), are presented in table 5.6. Bayes factors are calculated using equation

5.16 where M_1 is the logit model and model M_2 is the probit model.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	8.3852	8.7658	0.8796	0.1739
Gamma(5,15)	2.2889	2.3476	2.5796	2.1232
Uniform(0,1)	0.3354	0.6439	0.4218	3.0112

Table 5.6: $P(N|M_P)/P(N|M_L)$

The Bayes factors presented in table 5.6 show that the probit model is similar to the logit model for the three prior distributions and there is no evidence to suggest that the probit model is a better fit to the data than the logit model. When a Gamma(1,10) prior is used, for the first 100000 iterations of the MCMC the Bayes factor shows a preference for the probit model but for the second 100000 iterations there is a preference for the logit model. Using a Gamma(5,15) prior the Bayes factors remain fairly constant over the iterations, showing a slight preference for the probit model. A uniform prior shows a slight preference for the logit model until the final set of 50000 iterations which changes the preference to the probit model. These results show no substantial difference between the logit and probit models when using the data set where the comparison between death and the health state $(5, 2, 1, 1, 1)$ is removed.

5.5.2 Bayes factors when one death comparison altered

In the original data set out of the 42 people that compared the health states $(5, 2, 1, 1, 1)$ and death, one preferred death and 41 preferred health state $(5, 2, 1, 1, 1)$. This section presents the results when the data set is changed to 0 respondents preferring death to health state $(5, 2, 1, 1, 1)$.

Comparing the logit model with the benchmark model

Bayes factors supporting evidence in favour of the logit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1) are presented in table 5.7.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	4.3728×10^{121}	6.0722×10^{121}	6.3994×10^{121}	3.8794×10^{121}
Gamma(5,15)	4.8910×10^{116}	6.0928×10^{116}	6.7113×10^{116}	8.0730×10^{116}
Uniform(0,1)	7.8064×10^{121}	5.0632×10^{120}	6.4374×10^{120}	8.2416×10^{120}

Table 5.7: $P(N|M_L)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the logit model over the benchmark model. The Bayes factor values are larger than those calculated using the initial data in table 5.1 and for those where the comparison between death and health state $(5, 2, 1, 1, 1)$ is removed, in table 5.4.

Comparing the probit model with the benchmark model

Bayes factors supporting evidence in favour of the probit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1) are presented in table 5.8.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	8.2386×10^{121}	4.3574×10^{121}	3.5328×10^{121}	3.7907×10^{121}
Gamma(5,15)	2.3595×10^{116}	4.0827×10^{116}	5.7959×10^{116}	7.2835×10^{116}
Uniform(0,1)	2.1493×10^{121}	3.1621×10^{121}	2.7776×10^{121}	2.0021×10^{121}

Table 5.8: $P(N|M_P)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the probit model over the random model. The Bayes factor values are similar to those comparing the logit model and the benchmark model in this section, presented in table 5.7. The Bayes factors in 5.8 are larger than those calculated using both the initial data in table 5.2 and for those where the comparison between death and health state $(5, 2, 1, 1, 1)$ is removed, in table 5.4. The largest difference is between the table 5.8 and table 5.5.

Comparing the probit model with the logit model

Bayes factors to evaluate evidence in favour of the probit model over the logit model using the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1), are presented in table 5.9. Bayes factors are calculated using equation (5.16) where M_1 is the logit model and model M_2 is the probit model.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	1.8840	0.7176	0.5521	0.9771
Gamma(5,15)	0.4824	0.6701	0.8636	0.9022
Uniform(0,1)	0.2753	6.2453	4.3149	2.4293

Table 5.9: $P(N|M_P)/P(N|M_L)$

As with the Bayes factors in table 5.6, table 5.9 show there is no evidence of a difference between the logit and probit models for any of the three priors used.

5.5.3 Bayes factors after assuming negative error

This section presents the Bayes factor results when the error term in the utility model is assumed to be negative. The reasons for considering this approach are explained in the discussion in section 5.4.4. The utility, U_{ij}, of health state $\boldsymbol{x}_{ij}$ is defined to be

$$U_{ij} = g(\boldsymbol{x}_{ij}) - \varepsilon_{ij}. \tag{5.22}$$

Comparing the logit model with the benchmark model

Bayes factors supporting evidence in favour of the logit model against the benchmark model for each of the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1) are presented in table 5.10.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	3.1992×10^{119}	3.6891×10^{119}	4.5210×10^{119}	5.8477×10^{119}
Gamma(5,15)	3.1652×10^{115}	5.5059×10^{115}	7.3905×10^{115}	8.1985×10^{115}
Uniform(0,1)	1.7520×10^{120}	1.6864×10^{120}	1.5467×10^{120}	1.9730×10^{119}

Table 5.10: $P(N|M_L)/P(N|M_B)$

The Bayes factors show decisive evidence in favour of the logit model over the random model for all three prior distributions. The Bayes factor values are larger than those calculated using the initial data in table 5.1 and for those where the comparison between death and health state $(5, 2, 1, 1, 1)$ is removed, in table 5.4. However, they are a little smaller than those calculated in table 5.7, where one comparison is altered.

Comparing the probit model with the benchmark model

The normal distribution is symmetric and therefore when a negative error is assumed the likelihood for the probit model will be identical to when a positive error is assumed. The Bayes factors will be the same as table 5.2 in section 5.4.

Comparing the probit model with the logit model

Bayes factors to evaluate evidence in favour of the probit model over the logit model using the three prior distributions: Gamma(1,10), Gamma(5,15) and Uniform(0,1), are presented in table 5.11. Bayes factors are calculated using equation 5.16 where M_1 is the logit model and model M_2 is the probit model.

Prior Distribution	Number of Iterations			
	50000	100000	150000	200000
Gamma(1,10)	0.0123	0.0047	0.0021	0.0020
Gamma(5,15)	0.0012	0.0003	0.0003	0.0004
Uniform(0,1)	0.0019	0.0003	0.0005	0.0041

Table 5.11: $P(N|M_P)/P(N|M_L)$

The Bayes factors in 5.11 show a preference for the logit model over the probit model for all three prior distributions. This is the opposite to the results using the initial data in table 5.3. However the Bayes factors in table 5.3 show more evidence in favour of the probit model than the Bayes factors in table 5.11 do in favour of the logit model.

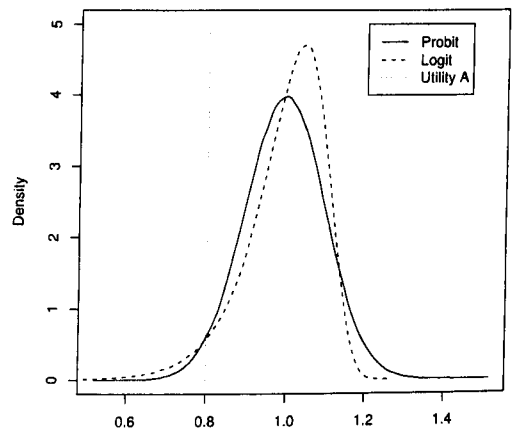


Figure 5.2: Probability distributions of health state (5,2,1,1,1)

Figure 5.2 presents the distributions of the health state $(5, 2, 1, 1, 1)$ using both the logit and probit models and assuming a negative error. As with figure 5.1,

utility A represents the utility the individual who prefers death has for health state $(5,2,1,1,1)$. In figure 5.2, the density to the left of utility A is greater for the logit than the probit model although the difference does not appear to be as great as the difference in figure 5.1.

5.6 Bayes Factors comparing models with reduced models

In section 5.4 Bayes factors are calculated using the original data set where each model has variables related to each of the five attributes in the classification system. This section considers five reduced models for the original data set, where each model has the variables for one attribute removed. In the full model the utility individual i has for health state $\boldsymbol{x}_{ij}$ is given by

$$U_{ij} = 1 - \boldsymbol{\theta}^T \boldsymbol{x}_{ij} + \varepsilon_{ij}, \tag{5.23}$$

where $\boldsymbol{\theta}$ is a vector of 21 unknown parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_{21})$. Each of the first 20 parameters represent the decrease in utility associated with the increase of an attribute by one level. The parameter θ_{21} represents the decrease in utility from perfect health to the health state death. There are four parameters associated with each attribute. For example, parameters $\theta_1, \theta_2, \theta_3, \theta_4$ are associated with incrementally increasing the level of the attribute concern about asthma. The model where the attribute concern about asthma is removed is defined as in equation (5.23) where $\theta_i = 0 \quad \forall i = 1, ..., 4$. Similar models are created when removing each of the other four attributes.

Bayes factors are then calculated to compare each of the reduced models with the full model in section 5.4. The reduced models are fitted for the logit model using both the Gamma(1,10) and the Uniform(0,1) prior distributions. This approach can be used to test for individual attribute effects in the full model and

each Bayes factor value represents the importance of a particular attribute to the model.

5.6.1 Gamma(1,10) Prior and logit Likelihood

Table 5.12 shows the Bayes factors comparing the full model to each of the reduced models for the logit likelihood and Gamma(1,10) prior distribution. Bayes factors are calculated using equation 5.16 where M_1 is the reduced model and model M_2 is the full model.

Attribute Removed	Number of Iterations			
	50000	100000	150000	200000
Concern	2.0595×10^7	1.0674×10^8	1.3038×10^8	1.3698×10^8
Breath	1.2171×10^9	4.5624×10^{10}	5.0117×10^{10}	4.6769×10^{10}
Weather	8.2599×10^8	7.8981×10^8	7.5869×10^8	7.0386×10^8
Sleep	7.7900×10^5	1.2512×10^6	1.7995×10^6	2.6108×10^6
Activities	3.6834×10^{37}	7.6792×10^{37}	1.4863×10^{38}	1.6600×10^{38}

Table 5.12: Bayes Factors comparing logit and Reduced logit model using gamma(1,10) prior

The Bayes factors show a preference for the full model over all the reduced models. The size of the Bayes factors show substantial evidence in favour of the full model. The largest Bayes factor values are calculated when comparing the full model to the model with the attribute activity limitations removed. This shows the greatest change in the model is when this attribute is removed and results in the largest effect on the utilities for the health states in the AQL-5D classification system. The Bayes factors when removing the other attributes are fairly similar in comparison and suggests these all have a similar effect on the full utility model. The Bayes factors remain fairly constant over the number of iterations.

The effect of each attribute is also shown by the parameter estimates for the logit model in chapter 4. The largest parameter estimate is associated with a change from level 3 to level 4 for the attribute activity limitations. The sum of the four parameters associated with an attribute has the largest value for the attribute activity limitations. The results agree with the Bayes factors, showing that the attribute activity limitations has the most value in the utility model.

5.6.2 Uniform(0,1) Prior and logit Likelihood

Table 5.13 shows the Bayes factors comparing the full model to each of the reduced models for the logit likelihood and Uniform(0,1) prior distribution. Bayes factors are calculated using equation 5.16 where M_1 is the reduced model and model M_2 is the full model.

Attribute Removed	Number of Iterations			
	50000	100000	150000	200000
Concern	1.5780×10^{8}	7.5206×10^{7}	8.2646×10^{7}	1.4373×10^{8}
Breath	1.6446×10^{10}	1.5675×10^{10}	2.1299×10^{10}	2.8052×10^{10}
Weather	3.1849×10^{8}	1.6085×10^{8}	2.7547×10^{8}	3.1942×10^{8}
Sleep	3.2608×10^{8}	9.2856×10^{7}	7.5922×10^{7}	7.1068×10^{7}
Activities	9.4654×10^{38}	6.1237×10^{39}	5.8273×10^{39}	3.0656×10^{40}

Table 5.13: Bayes Factors comparing logit and Reduced logit model with Uniform(0,1) prior

The results in table 5.13 show a similar pattern to the Bayes factors for the Gamma(1,10) prior distribution in table 5.12. The largest Bayes factors are when the full model is compared to the model with the attribute activity limitations removed. All other Bayes factors are similar in comparison. The Bayes factors

remain fairly constant over the number of iterations and are slightly larger than those for the Gamma(1,10) prior, but not significantly.

5.7 Conclusion

The Bayes factors show that both the logit model and probit model are suitable for the AQL-5D discrete choice data when compared with the benchmark model. The inclusion of an unusual observation in the comparison between death and the health state $(5, 2, 1, 1, 1)$ resulted in an unexpected large Bayes factor in favour of the probit model when compared with the logit model. When the observation is removed the Bayes factors were as expected, showing no difference between the two models. This problem showed that the evidence a single Bayes factor presents should only be considered when used in comparison with other Bayes factors.

Each attribute in the AQL-5D classification system helps explain the data and is important in defining utilities. The attribute activity limitations has been shown to have the greatest impact on the utility.

Chapter 6

Random effects Model

6.1 Introduction

In this chapter we consider including a random effect in the logit model for defining utility of AQL-5D health states. A review of different random effects models is undertaken, followed by a definition of a random effects logit model and how this relates to the mixed logit model. Results are presented using both maximum likelihood estimation and Bayesian methods. Several starting values are used in both methods and an investigation into differences in the results is undertaken.

6.2 Motivation of Random effects model

In chapters 4 and 5 the utility individual i has for health state $\boldsymbol{x}_{ij}$ is defined as

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \varepsilon_{ij}, \quad j = 1, ...J, \tag{6.1}$$

where $g(\boldsymbol{x}_{ij})$ is a function of $\boldsymbol{x}_{ij}$ with unknown parameters and represents the population mean utility. The error, ε_{ij}, represents the individual's variation in

preference from the population mean utility. Each error is assumed to be independently and identically distributed.

The assumption of independence is questionable. Consider two similar health states, $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$, that are defined using the AQL-5D classification system. Suppose $\boldsymbol{x}_{ia}$ describes the health state where concern about asthma is experienced some of the time, shortness of breath as a result of asthma is experienced some of the time, asthma symptoms as a result of pollution are experienced some of the time, asthma interferes with sleep some of the time and there is moderate limitation in every activity done. In numerical form this is written as $(3, 3, 3, 3, 3)$. Suppose health state $\boldsymbol{x}_{ib}$ is the same as $\boldsymbol{x}_{ia}$ for the first four attributes but for the fifth there is extreme limitations in every activity done. In numerical form this is written $(3, 3, 3, 3, 4)$. If the utility individual i has for health state $\boldsymbol{x}_{ia}$ is less than the population mean utility for health state $\boldsymbol{x}_{ia}$, then we might expect that the utility individual i has for health state $\boldsymbol{x}_{ib}$ is also less than the population mean utility. Consequently, we would judge the errors associated with the two health states, $\boldsymbol{x}_{ia}$ and $\boldsymbol{x}_{ib}$, to be correlated, and regard the assumption of independent and identically distributed errors to be not valid.

Correlation can be accounted for in the model by including a random effect. In this chapter two random effects models are reviewed, the additive model and the multiplicative model. The logit model is fitted to the AQL-5D data set described in chapter 4, where the utility is defined using a multiplicative random effect. The model is then fitted using both maximum likelihood and Bayesian inference.

6.3 Types of random effects models

A random effect can be additive or multiplicative. If an additive random effect is considered, the utility in equation (6.1) is now written as

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \alpha_i I\{\boldsymbol{x}_{ij} \neq \text{death or perfect health}\} + \varepsilon_{ij}, \quad j = 1, ...J, \quad (6.2)$$

where α_i is the random effect for individual i, and $I\{\boldsymbol{x}_{ij} \neq$ death or perfect health$\} =$ 0 if $\boldsymbol{x}_{ij} =$ death or perfect health and 1 otherwise. The random effect is fixed for each individual i and represents the mean deviation of individual i's utilities from the population mean utilities. The error, ε_{ij}, represents additional departures from the mean utility and these can now be considered to be independent and identically distributed. If the health state $\boldsymbol{x}_{ij}$ is perfect health then $U_{ij} = 1$. Therefore $g(\boldsymbol{x}_{ij}) = 1$ and $\varepsilon_{ij} = 0$. If the health state $\boldsymbol{x}_{ij}$ is death then $U_{ij} = 0$. Therefore $g(\boldsymbol{x}_{ij}) = 0$ and $\varepsilon_{ij} = 0$. The condition $I\{\boldsymbol{x}_{ij} \neq$ death or perfect health$\}$ ensures that the random effects are not included for the utility of either death or perfect health. Figure 6.1 shows a graphical representation of an additive random effects model. Line A presents the mean of 48 health states plotted in decreasing order. Line B presents the utilities of an individual with random effects $\alpha_i = -0.1$. The error, ε_{ij}, represents the additional departures of individual i's utilities from line B.

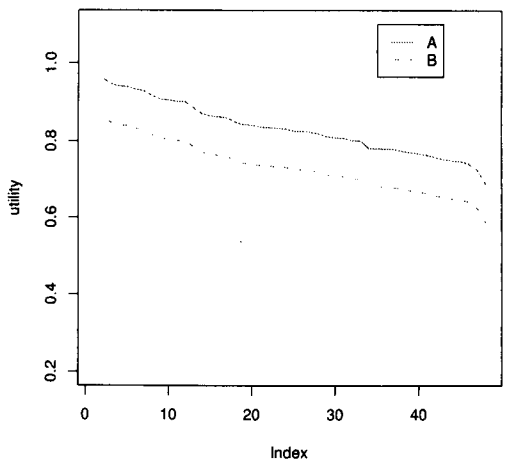


Figure 6.1: Ordered utility of 48 health states using an additive random effects model where line A is the mean utility and line B is the utility of an individual with random effect $\alpha = -0.1$

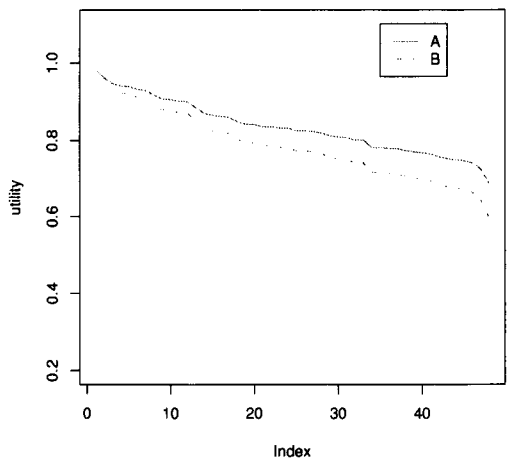


Figure 6.2: Ordered utility of 48 health states using a multiplicative random effects model where line A is the mean utility and line B is the utility of an individual with random effect $\alpha = 1.3$

The variability in respondents' utilities for a health state close to death might be expected to be greater than the variability in respondent's utilities for a health state closer to perfect health. A health state utility cannot be greater than 1, which decreases the variability of the utility of a health state close to perfect health. The utility of death is 0. A health state can have a utility of less than 0, so the utility of any health state close to death has greater variability than the utility of a health state close to perfect health. It might be considered unlikely that an individual's utility for a health state similar to perfect health is significantly less than 1. A health state close to death would not be expected to have a utility close to 0 for every individual. One individual might consider all the health states in a classification system significantly better than death, and

therefore every health state has a utility close to 1. Another individual might consider some health states to be worse than death. An additive random effect assumes that the variability in respondents utilities is the same for each health state, which might not be considered a reasonable assumption for health state utilities. Therefore an additive random effect is not appropriate and another type of random effect should be considered.

An example of a random effect that allows different levels of variation is a multiplicative random effect. Let α_i be the multiplicative random effect associated with individual i. Define

$$d_{ij} = \begin{cases} 0 & \text{if health state } x_{ij} \text{ is death or perfect health,} \\ 1 & \text{otherwise.} \end{cases} \tag{6.3}$$

A multiplicative random effects model, as defined in Kharroubi et al. (2005), is

$$U_{ij} = 1 - (\alpha_i)^{d_{ij}}(1 - g(\boldsymbol{x}_{ij})) + \varepsilon_{ij}, \quad j = 1, ..., K. \tag{6.4}$$

If $\alpha_i < 1$, individual i's utility is greater than the population mean utility. If $\alpha_i > 1$, individual i's utility is less than the population mean utility. Respondents with the largest values of α_i are the most likely to value a health state as worse than death. The error, ε_{ij}, represents additional departures from the mean utility for individual i. An example of a multiplicative random effect is presented in Figure 6.2. Line A represents the population mean utilities of 48 health states in decreasing order. Line B represents the utilities for individual i where the random effect is $\alpha_i = 1.3$. The health states in line B that are close to perfect health are very similar to the population mean. As the health states get closer to the health state death, the difference between the population mean utility and individual i's utility increases.

6.3.1 Relationship of multiplicative random effects model with Mixed Logit Model

Suppose a multiplicative random effects is assumed and the utility, U_{ij}, of health state $\boldsymbol{x}_{ij}$ is defined as

$$U_{ij} = \alpha_i \boldsymbol{\theta}^T \boldsymbol{x}_{ij} + \varepsilon_{ij}, \tag{6.5}$$

where α_i is the random effect for individual i. As defined in equation (3.53), the utility using a mixed logit model is written as

$$U_{ij} = \boldsymbol{\theta}_i \boldsymbol{x}_{ij} + \varepsilon_{ij}, \tag{6.6}$$

where $\boldsymbol{\theta}_i$ is the vector of parameters for individual i. Equation (6.5) can be written as equation (6.6) by letting $\boldsymbol{\theta}_i = \alpha_i \boldsymbol{\theta}$. However, equation (6.6) can only be written as equation (6.5) if values of α_i and $\boldsymbol{\theta}$ are found where $\boldsymbol{\theta}_i = \alpha_i \boldsymbol{\theta}$.

6.4 Logit Model

If the errors in equation 6.4 are assumed to have a type 1 extreme value distribution then the probability associated with choosing a health state is given by the logit model. As described in chapter 4, each respondent in the AQLQ study assesses 8 pairwise comparisons. Let $B_j = \{\boldsymbol{x}_{ij1}, \boldsymbol{x}_{ij2}\}$ be the set containing the j^{th} pair of health states to be compared by individual i. Define

$$n_{ij} = \begin{cases} 1 & \text{if individual } i \text{ prefers } \boldsymbol{x}_{ij1} \text{ to } \boldsymbol{x}_{ij2} \\ 0 & \text{if individual } i \text{ prefers } \boldsymbol{x}_{ij2} \text{ to } \boldsymbol{x}_{ij1} \end{cases} \tag{6.7}$$

Some sets will include the health state death. For these sets it is assumed that $\boldsymbol{x}_{ij2}$ is the health state death. Define $M = \{m_{ij}, i = 1, ...I, j = 1, ...J_i\}$ with

$$m_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{x}_{ij2} \text{ is the health state death} \\ 0 & \text{if } \boldsymbol{x}_{ij2} \text{ is not the health state death} \end{cases} \tag{6.8}$$

and $N = \{n_{ij}, \quad i = 1 \ldots I, \quad j = 1 \ldots J_i\}$. The likelihood function for the individuals' choices is

$$P(N|\boldsymbol{\theta}, \sigma^2, M) = \prod_{i=1}^{I} \prod_{j=1}^{J_i} \left\{ \left[\frac{\exp\left(\frac{h(\boldsymbol{x}_{ij1})}{\sigma}\right)}{\exp\left(\frac{h(\boldsymbol{x}_{ij1})}{\sigma}\right) + \exp\left(\frac{h(\boldsymbol{x}_{ij2})}{\sigma}\right)} \right]^{n_{ij} \times (1 - m_{ij})} \right.$$

$$\times \left[1 - \frac{\exp\left(\frac{h(\boldsymbol{x}_{ij1})}{\sigma}\right)}{\exp\left(\frac{h(\boldsymbol{x}_{ij1})}{\sigma}\right) + \exp\left(\frac{h(\boldsymbol{x}_{ij2})}{\sigma}\right)} \right]^{(1 - n_{ij}) \times (1 - m_{ij})}$$

$$\times \left[1 - \exp\left(- \exp\left(\frac{h(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right]^{n_{ij} \times m_{ij}}$$

$$\left. \times \left[\exp\left(- \exp\left(\frac{h(\boldsymbol{x}_{ij1}) - 0.5722\sigma}{\sigma} \right) \right) \right]^{(1 - n_{ij}) \times m_{ij}} \right\} \tag{6.9}$$

where $h(\boldsymbol{x}_{ijk}) = 1 - \alpha_i^{m_{ij}}(1 - g(\boldsymbol{x}_{ijk}))$, $\quad k = 1, 2$, and $g(\boldsymbol{x}_{ijk})$ is assumed to be the linear model described in chapter 4 where $g(\boldsymbol{x}_{ijk}) = 1 - \boldsymbol{x}_{ijk}^T \boldsymbol{\theta}$.

6.5 Maximum Likelihood Results

This section presents values for $\boldsymbol{\theta}$, σ and summary statistics of the elements of $\boldsymbol{\alpha}$ estimated using maximum likelihood. The results are derived using numerical methods, implemented using the *nlm* command in **R**, which minimises a function using a Newton-type algorithm. The results after a maximum likelihood estima-

tion can be influenced by the length of time the maximisation algorithm is run and also the starting values. Starting values are initial guess of the parameter values but if the starting values are too far from the true parameter values then the derived maximum could be a local maximum instead of a global maximum. It is therefore important to run the maximum likelihood algorithm using several different starting values.

Maximum likelihood values are derived for the parameters $\boldsymbol{\theta}$, σ and $\boldsymbol{\alpha}$ using several different sets of starting values. The results are presented in table ?? of the appendix, by starting values for $\boldsymbol{\theta}$ and σ. The maximum likelihood are very variable and depend on the starting values used. For all starting values, the largest parameter estimate is for the attribute sleep at level 5. However the parameter estimate ranges between 0.0054 and 2.9750. The starting value for $\boldsymbol{\alpha}$ appears to have the greatest effect on the results. Small starting values for $\boldsymbol{\alpha}$ result in a small estimated mean and variance of $\boldsymbol{\alpha}$ and larger estimates of $\boldsymbol{\theta}$ and σ. Larger starting values for α result in a larger mean and variance of the random effect estimates and smaller estimates of $\boldsymbol{\theta}$ and σ.

6.6 Bayesian Inference

We now consider Bayesian inference for the model in equation (6.3). Prior distributions are required for $\boldsymbol{\theta}$, the scale parameter σ and the vector $\boldsymbol{\alpha}$ of 168 random effects of the 168 individuals in the study.

6.6.1 Prior Distributions

Each parameter $\theta_i, \quad i = 1, ...20$ represents the decrease in utility associated with the level of an attribute. As no health state can have a utility greater than 1, a convenient prior distribution for each parameter θ_i is the gamma distribution

with suitable shape and rate values. The same prior distribution is also considered for the scale parameter, as explained in chapter 4.

The random effects are always positive and therefore the prior distribution assumed must ensure $\alpha_i \geq 0$. For computational convenience a lognormal prior distribution is therefore assumed for each random effect, $\log(\alpha_i) \sim N(\mu, \tau)$. Initially, the parameter μ is assumed to be $\mu = 0$. Therefore the random effects have a median of 1. The variance, τ, is unknown and therefore a prior distribution is assumed. A computationally convenient prior is the inverse gamma distribution, which has pdf

$$f(\tau) = \frac{b^a}{(a-1)!}\tau^{-a-1}\exp\left(\frac{-b}{\tau}\right), \tag{6.10}$$

where a is the shape parameter and b is the scale parameter.

Prior distributions are considered where τ is less than 1 as it is assumed that the random effects would not deviate significantly from 1. Three prior distributions are compared: the Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5). These are presented graphically in figures (6.3), (6.4) and (6.5). Table 6.2 presents the 5% and 95% percentiles of each prior distribution for τ. For each of these two values of the τ parameter the 5% and 95% percentiles of the corresponding random effects distribution are also presented. This shows that the range of possible values for the random effects does not vary significantly from 1.

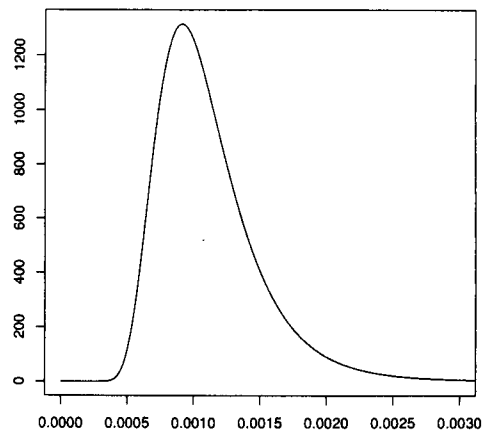


Figure 6.3: Inverse-Gamma(10,0.01) prior

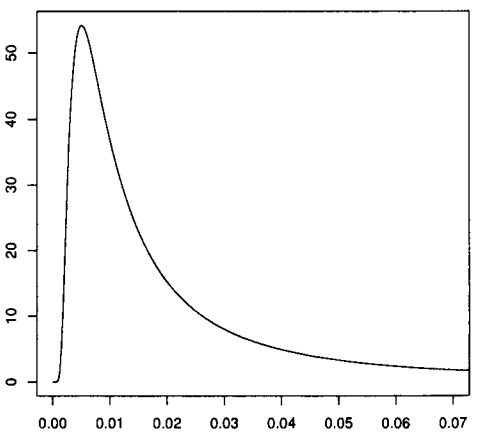


Figure 6.4: Inverse-Gamma(0,0.01) prior

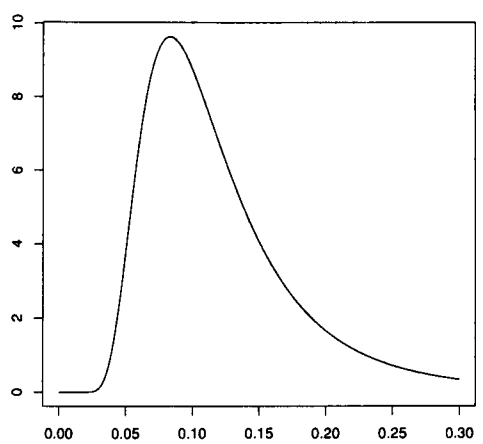


Figure 6.5: Inverse-Gamma(5,0.5) prior

Shape	Scale	$\tau_{0.05}$	$\tau_{0.95}$	$\alpha_{0.05}(\tau_{0.05})$	$\alpha_{0.95}(\tau_{0.05})$	$\alpha_{0.05}(\tau_{0.95})$	$\alpha_{0.95}(\tau_{0.95})$
10	0.01	0.006	0.0018	0.999	1.001	0.997	1.003
1	0.01	0.0033	0.195	0.995	1.005	0.726	1.378
5	0.5	0.055	0.254	0.914	1.095	0.658	1.519

Table 6.2: 5th and 95th percentiles of each Inverse-Gamma prior distribution on τ and the corresponding 5th and 95th percentiles for the distribution of α

6.6.2 Posterior Inference

Posterior distributions are required for the parameters $\theta_1, ..., \theta_{20}$ and σ in the logit model and can be inferred by simulating from the joint distribution $f(\theta_1, ..., \theta_{20}, \sigma, \boldsymbol{\alpha}, \tau)$. MCMC simulation samples vectors of values $\boldsymbol{\pi} = (\theta_1, ..., \theta_{20}, \sigma, \boldsymbol{\alpha}, \tau)$. The sequence of vectors $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, ...$ is a Markov Chain. The stationary distribution of the Markov chain is the required posterior distribution.

Let the state of the Markov chain at time t be

$$\boldsymbol{\pi}^{(t)} = (\theta_1^{(t)}, ..., \theta_{20}^{(t)}, \sigma^{(t)}, \boldsymbol{\alpha}^{(t)}, \tau^{(t)}). \tag{6.11}$$

To update this vector to the state at time $t+1$,

$$\boldsymbol{\pi}^{(t+1)} = (\theta_1^{(t+1)}, ..., \theta_{20}^{(t+1)}, \sigma^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \tau^{(t+1)}), \tag{6.12}$$

a new value is generated for each of the parameters. The process of updating the parameters at time t, $\boldsymbol{\pi}^{(t)}$, to the parameters at time $t+1$, $\boldsymbol{\pi}^{(t)}$, is called an iteration.

Gibbs sampling is used to simulate values from the conditional posterior distributions of each parameter. The 168 random effects parameters are updated

as a vector, $\boldsymbol{\alpha}$, to decrease the time taken to run the program. For this model, one iteration of the Gibbs sampler has 22 steps. Given data N, to update the parameter vector at time t, to the vector at time t+1,

$$
\begin{aligned}
&\text{Sample} \quad \theta_1^{(t+1)} \quad \text{from} \quad f(\theta_1|\theta_2^{(t)},...,\theta_{20}^{(t)},\sigma^{(t)},\alpha_1^{(t)},...,\alpha_{168}^{(t)},\tau^{(t)},N)\\
&\text{Sample} \quad \theta_2^{(t+1)} \quad \text{from} \quad f(\theta_2|\theta_1^{(t+1)},\theta_3^{(t)},...,\theta_{20}^{(t)},\alpha_1^{(t)},...,\alpha_{168}^{(t)},\tau^{(t)},N)\\
&\quad .\\
&\quad .\\
&\quad .\\
&\text{Sample} \quad \boldsymbol{\alpha}^{(t+1)} \quad \text{from} \quad f(\boldsymbol{\alpha}|\theta_1^{(t+1)},...,\theta_{20}^{(t+1)},\tau^{(t)},N)\\
&\text{Sample} \quad \tau^{(t+1)} \quad \text{from} \quad f(\tau|\theta_1^{(t+1)},...,\theta_{20}^{(t+1)},\boldsymbol{\alpha}^{(t+1)},N).
\end{aligned}
\tag{6.13}
$$

The only conditional distribution that can be derived analytically is for the variance parameter τ. The conditional posterior distribution for the parameter τ, $f(\tau|\boldsymbol{\theta},\sigma,\boldsymbol{\alpha},N)$, can be written as $f(\tau|\boldsymbol{\alpha})$, as the value of parameter τ, given $\boldsymbol{\alpha}$, is conditionally independent of $\boldsymbol{\theta}$, σ and N. The conditional posterior distribution, $f(\tau|\boldsymbol{\alpha})$, is derived using

$$f(\tau|\boldsymbol{\alpha}) \propto f(\boldsymbol{\alpha}|\tau)f(\tau). \tag{6.14}$$

If α_i has a lognormal distribution with parameters μ and τ, then $\varphi_i = \log(\alpha_i)$ has a normal distribution with mean μ and variance τ. Equation (6.14) can be written as

$$f(\tau|\boldsymbol{\alpha}) = f(\tau|\boldsymbol{\varphi}) \propto f(\boldsymbol{\varphi}|\tau)f(\tau). \tag{6.15}$$

If φ_i has a normal distribution and τ has an inverse-gamma distribution, then the conditional posterior distribution of τ is

$$f(\tau|\boldsymbol{\varphi}) \propto \prod_{i=1}^{I} \frac{1}{(2\pi\tau)^{1/2}} \exp\left(-\frac{1}{2}\frac{(\varphi_i-\mu)^2}{\tau}\right)$$

$$\times\, \tau^{-(a+1)} \exp\left(\frac{-b}{\tau}\right)$$

$$= \frac{1}{(2\pi)^{1/2}\quad \tau^{I/2}} \exp\left(-\frac{1}{2}\frac{(\sum_{i=1}^{I}(\varphi_i-\mu)^2}{\tau}\right) \tag{6.16}$$

$$\times\, \tau^{-(a+1)} \exp\left(\frac{-b}{\tau}\right)$$

$$= \frac{1}{(2\pi)^{1/2}} \tau^{-(a+1+I/2)} \exp\left(-\frac{1}{\tau}\left(b+\frac{(\sum_{i=1}^{I}\varphi_i-\mu)^2}{2}\right)\right).$$

Therefore the conditional posterior distribution, $\tau|\boldsymbol{\varphi}$, has an Inverse-Gamma distribution with the following parameters,

$$Inv-Gamma\left(a+\frac{I}{2},\quad b+\frac{\sum_{i=1}^{I}(\varphi_i-\mu)^2}{2}\right),$$

where I is the number of respondents. After substitution of φ_i and I, the conditional distribution, $\tau|\boldsymbol{\alpha}$, has the distribution

$$Inv-Gamma\left(a+84,\quad b+\frac{\sum_{i=1}^{168}(\log(\alpha_i)-\mu)^2}{2}\right).$$

The conditional distributions of the other parameters can be simulated from using the Metropolis-Hastings algorithm, as defined in chapter 4. Each parameter $\boldsymbol{\theta}$, σ and $\boldsymbol{\alpha}$ must be positive. A suitable proposal distribution is the log-normal distribution. The proposal distributions for each of the parameters are written as

$$q_i(\theta|\theta_{t,i})=\frac{1}{\theta\nu\sqrt{2\theta}}\exp\left[-\frac{1}{2}\left(\frac{\log\theta-\log\theta_{t,i}}{\nu}\right)^2\right], \tag{6.17}$$

$$q_i(\sigma|\sigma_{t,i})=\frac{1}{\sigma\nu\sqrt{2\sigma}}\exp\left[-\frac{1}{2}\left(\frac{\log\sigma-\log\sigma_{t,i}}{\nu}\right)^2\right], \tag{6.18}$$

$$q_i(\alpha|\alpha_{t,i})=$$

$$\frac{1}{\alpha\nu\sqrt{2\alpha}\exp\left[-\frac{1}{2}\left(\frac{\log\alpha-\log\alpha_{t,i}}{\nu}\right)^2\right]}. \tag{6.19}$$

6.7 MCMC Results

The posterior distribution can be affected by prior distribution, data and the starting values. The three prior distributions considered are described in section 6.6.1. Two starting values are used for each prior distribution and the results compared to investigate the effect of starting values on the results. The two sets of starting values and their corresponding results are presented in tables 6.3 and 6.5. The first set of starting values has a mean random effect close to 1, a small random effect variance and relatively large values of $\boldsymbol{\theta}$ and σ. The second set of starting values have a larger mean and variance for the random effects and smaller values for $\boldsymbol{\theta}$.

6.7.1 STARTING VALUES 1

An MCMC simulation was run with 120000 iterations. The first 20000 were discarded as burn-in. The starting values for the parameters $\boldsymbol{\theta}$ and σ are presented in table 6.3 and the distribution of starting values for the random effects is summarized in table 6.4. The starting value for the variance of the random effects, τ, depends on the prior distribution used. For the prior distributions Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) the starting values for τ are 0.001, 0.01 and 0.1 respectively.

Attribute	Concern	Breath	Weather	Sleep	Activities
Level 2	0.0493	0.0306	3.7357×10^{-10}	0.0931	0.0309
Level 3	0.0098	0.1011	0.0007	0.0005	0.0508
Level 4	0.0510	0.0003	0.0015	0.0450	0.0490
Level 5	0.0080	0.0006	0.0629	0.1819	0.0029
Scale			0.2341		

Table 6.3: Starting values for MCMC

Mean	Variance	Minimum	Maximum	Median
1.0786	0.0765	0.4428	1.8679	1.0600

Table 6.4: Summary of distribution of starting values for random effects

The posterior distribution of the variance of the random effects is compared for each of three prior distributions: Inverse Gamma(10,0.01), Inverse Gamma(1,0.01) and Inverse Gamma(5,0.5). Figure 6.6 presents the three posteriors using values of the variance selected between iteration 20001 and 70000 of the MCMC. Figure 6.7 presents the three posteriors using values of the variance selected between iteration 70001 and 100000 of the MCMC. The posterior distributions using Inverse-Gamma(10,0.01) and Inverse-Gamma(5,0.5) are the most similar in both graphs. The posterior using the prior distribution Inverse-Gamma(1,0.01) is further away from zero and appears to have a mean of approximately 5. The posterior distributions using an Inverse-Gamma(10,0.01) and Inverse-Gamma(5,0.5) priors have a mean of between 2 and 3. The posterior distributions are similar over both sets of 50000 iterations.

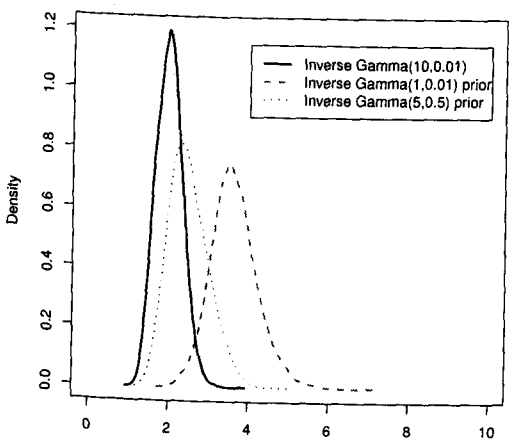


Figure 6.6: Posterior distributions by prior for iterations 20001 to 70000

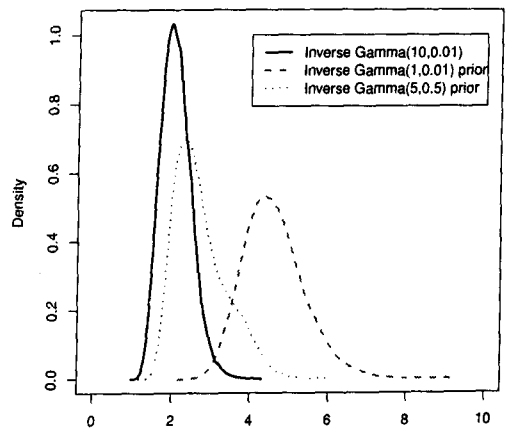


Figure 6.7: Posterior distributions by prior for iterations 700001 to 120000

The posterior distributions of the variance of the random effects are also compared with the prior distribution. Figure 6.8, 6.9 and 6.10 present the prior and posterior distribution of the variance parameter for each of the prior distribution Inverse Gamma(10,0.01), Inverse Gamma(1,0.01) and Inverse Gamma(5,0.5) respectively. Values selecting using iterations between 70001 and 120000 are used to plot the posterior distributions.The prior distributions are all distributed very close to zero. The three posterior distributions are distributed significantly further away from zero than the prior distributions. This implies the likelihood has the greatest influence on the posterior distribution.

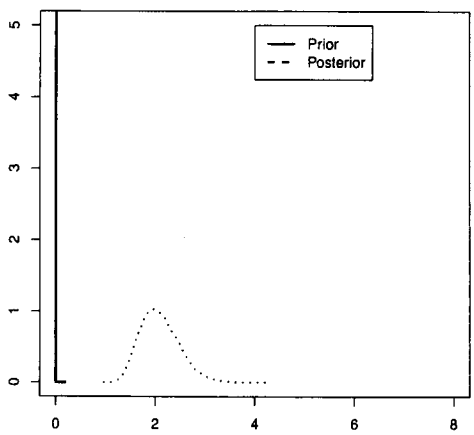


Figure 6.8: Prior and Posterior of Variance of Random effects for Inverse-Gamma(10,0.01) prior using iterations 70001 to 120000

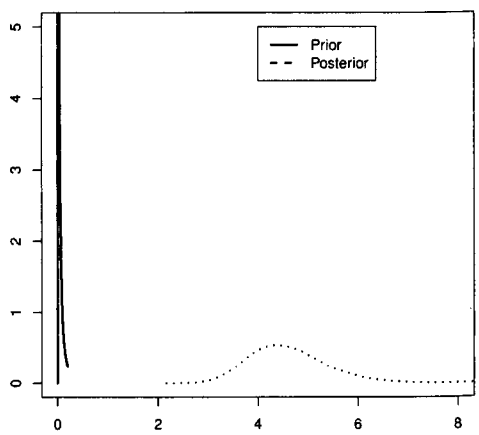


Figure 6.9: Prior and Posterior of Variance of Random effects for Inverse-Gamma(1,0.01) prior using iterations 70001 to 120000

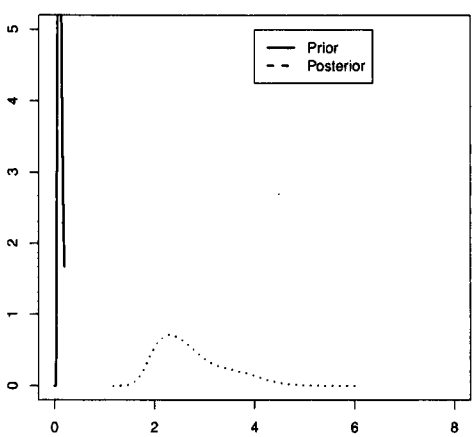


Figure 6.10: Prior and Posterior distribution for Inverse-Gamma(5,0.5) prior using iterations 70001 to 120000 ,

Convergence was assessed by plotting the value of the parameters at each iteration and calculating the acceptance rate. The iteration plots all show convergence and the acceptance rates are between 20% and 40% showing that the sampler is moving around the posterior distribution efficiently. The iteration plots also show consistency over both sets of 50000 iterations for all three prior distributions.

Using the last 50000 sample vectors of $\boldsymbol{\theta}_1, ...\boldsymbol{\theta}_{50000}$ and $\boldsymbol{\alpha}_1, ...\boldsymbol{\alpha}_{50000}$, utilities are calculated for the 48 health states in the sample survey. The mean and 95% posterior intervals are then derived for each health state. Figures 6.11, 6.12 and 6.13 present the mean and 95% posterior interval of the 48 health states using the Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) prior distributions for the variance of the random effects. The mean utilities using an Inverse-Gamma(1,0.01) and Inverse-Gamma(10,0.01) prior are similar. However, when using an Inverse-Gamma(5,0.5) prior the mean utilities and upper posterior limit are larger.

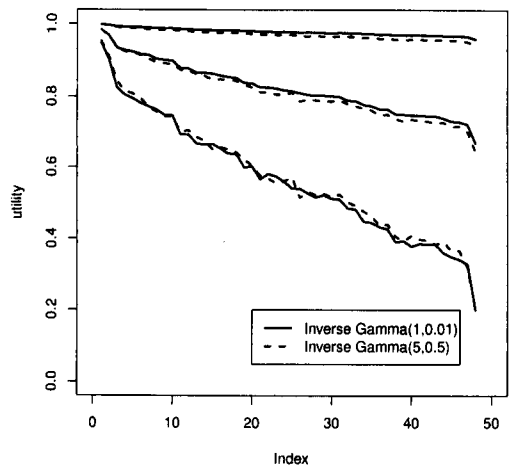


Figure 6.11: Mean utility and posterior interval of 48 health states using Inverse-Gamma(10,0.01) and Inverse(1,0.01) priors for variance of random effects for iteration 70001 to 120000

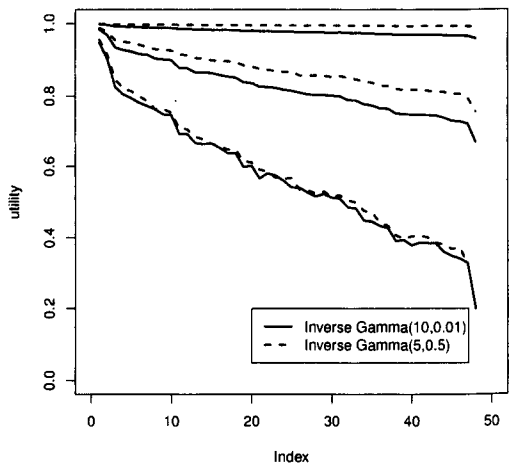


Figure 6.12: Mean utility and posterior interval of 48 health states using Inverse-Gamma(10,0.01) and Inverse(5,0.5) priors for variance of random effects using iteration 70001 to 120000

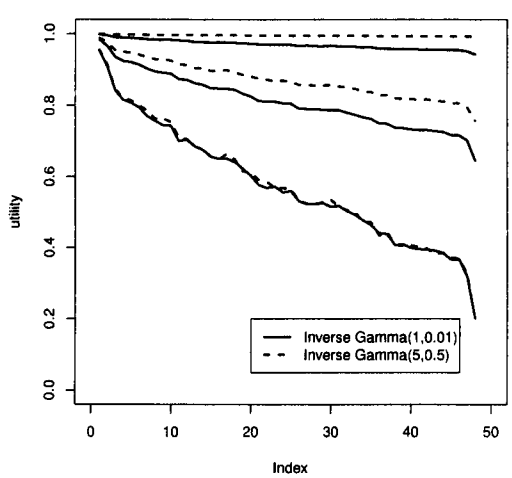


Figure 6.13: Mean utility and posterior interval of 48 health states using Inverse-Gamma(1,0.01) and Inverse(5,0.5) priors for variance of random effects for iteration 70001 to 120000

6.7.2 STARTING VALUES 2

A second MCMC was run with 120000 iterations. The first 20000 were discarded as burn-in. The starting values for each the parameters $\boldsymbol{\theta}$ and σ are shown in table 6.5 and the distribution of starting values for the random effects is summarised in table 6.6. The starting values for τ are 0.001, 0.01 and 0.1 for the Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) prior distributions respectively. For the prior distributions Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) the starting values for the variance τ are 0.001, 0.01 and 0.1 respectively.

Attribute	Concern	Breath	Weather	Sleep	Activities
Level 2	0.0033	6.5465×10^{-5}	1.8880×10^{-151}	0.0031	0.0023
Level 3	0.0003	0.0022	0.0001	2.9981×10^{-5}	0.0022
Level 4	0.0012	1.0167×10^{-11}	0.0013	1.4474×10^{-6}	0.0014
Level 5	7.9263×10^{-40}	0.0012	0.0027	0.0027	0.0014
Scale			0.1018		

Table 6.5: Starting values for MCMC

Mean	Variance	Minimum	Maximum	Median
27.7781	587.3725	2.3727×10^{-8}	157.7109	30.7537

Table 6.6: Summary of distribution of starting values for random effects

Figure 6.14 presents the three posteriors using values of the variance selected between iteration 20001 and 70000 of the MCMC. The posteriors show differences over the prior distribution. The posterior distribution derived using an Inverse-Gamma(10,0.01) prior is distributed between approximately 10 and 30. The posterior using an Inverse-Gamma(1,0.10) prior is distributed further away, be-

tween approximately 20 and 55. The posterior using an Inverse-Gamma(5,0.5) prior is bimodal and is distributed between approximately 5 and 30. Figure 6.15 presents the three posteriors using values of the variance selected between iteration 70001 and 100000 of the MCMC. For the Inverse-Gamma(10,0.01) and Inverse-Gamma(1,0.01) prior distributions, the posterior distributions are similar to those derived using the previous 50000 iterations in figure 6.14. The posterior distributions derived using an Inverse-Gamma(5,0.5) prior distribution now has a smaller variance, a larger mode and has become unimodal. Comparing the posterior distributions using each of the two starting values shows that the posterior distributions are distributed substantially further away from zero when using starting values 2 for the three prior distributions. For example, the posterior derived from an Inverse-Gamma(10,0.01) is distributed between approximately 1 and 4 for starting values 1 and is distributed between approximately 10 and 35 for starting values 2.

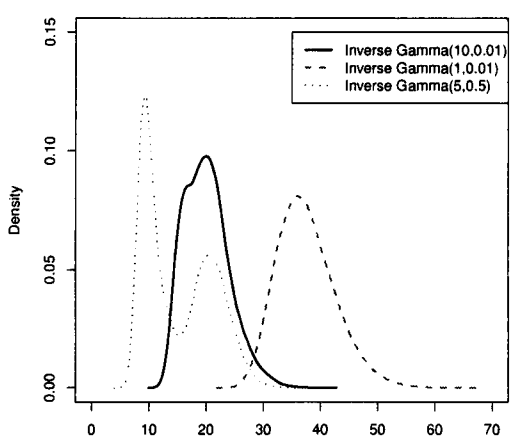


Figure 6.14: Posterior distributions by prior for iterations 20001-70000

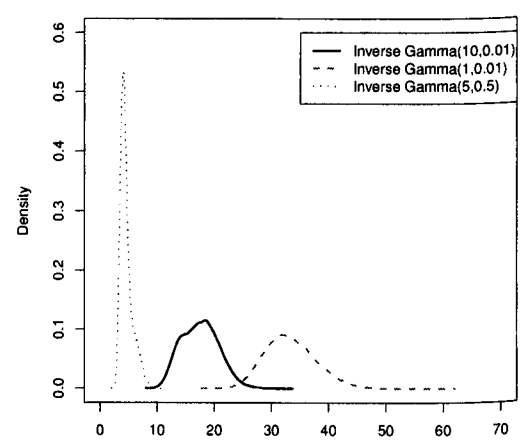


Figure 6.15: Posterior distributions by prior for iterations 700001 to 120000

As for starting values 1, the posterior distributions of the variance of the random effects are also compared with the prior distribution. Figure 6.16, 6.17 and 6.18 present the prior and posterior distribution of the variance parameter for the prior distributions Inverse Gamma(10,0.01), Inverse Gamma(1,0.01) and Inverse Gamma(5,0.5) respectively. Values selecting using iterations between 70001 and 120000 are used to plot the posterior distributions. The posterior distributions using starting values 2 show a greater change from the prior distributions than the posterior distributions for starting values 1, presented in figures 6.8, 6.9, and 6.10.

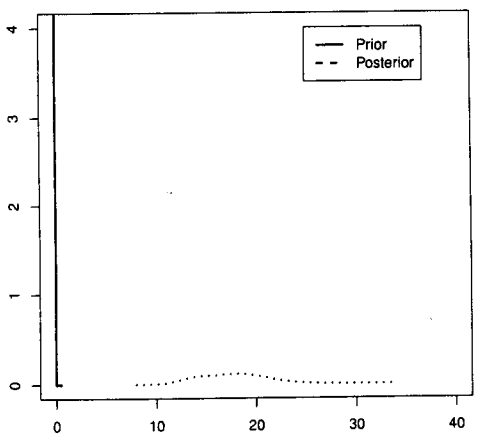


Figure 6.16: Prior and Posterior of Variance of Random effects for Inverse-Gamma(10,0.01) prior using iterations 70001 to 120000

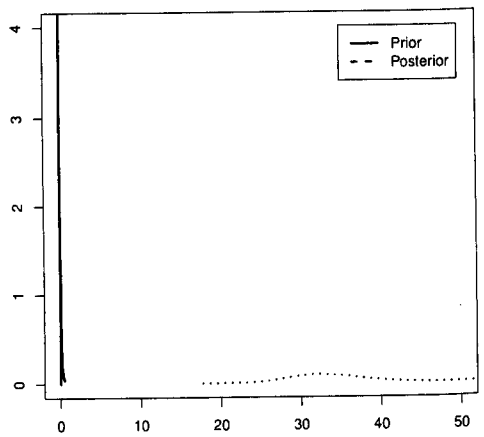


Figure 6.17: Prior and Posterior of Variance of Random effects when using Inverse-Gamma(1,0.01) prior using iterations 70001 to 120000

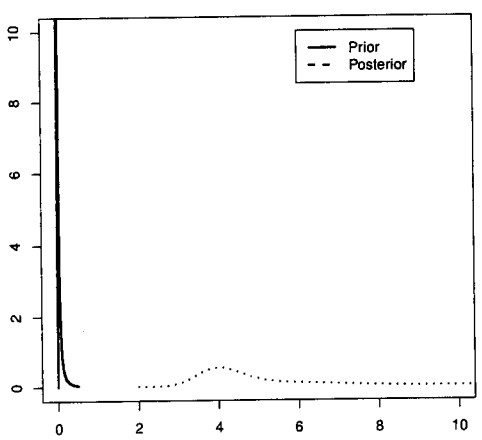
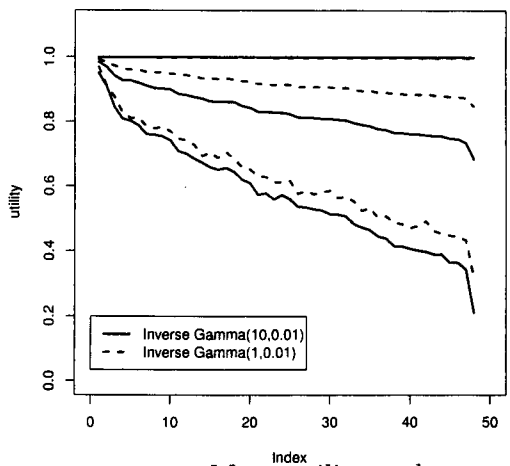


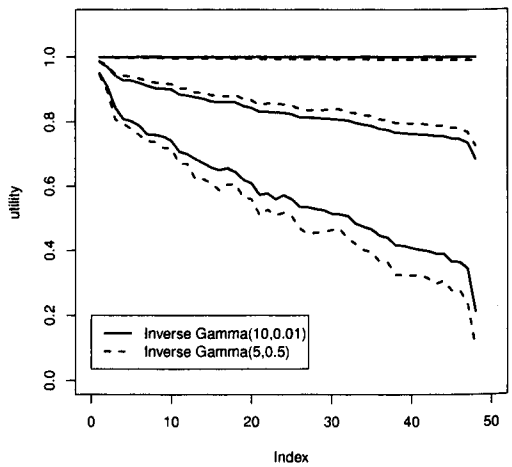
Figure 6.18: Prior and Posterior of Variance of Random effects when using Inverse-Gamma(5,0.5) prior using iterations 70001 to 120000

Convergence was assessed by plotting the value of the parameters at each iteration and calculating the acceptance rate. The iteration plots all show convergence and the acceptance rates are between 20% and 40% showing that the sampler is moving around the posterior distribution efficiently. The iteration plots also show consistency over both sets of 50000 iterations for all three prior distributions.

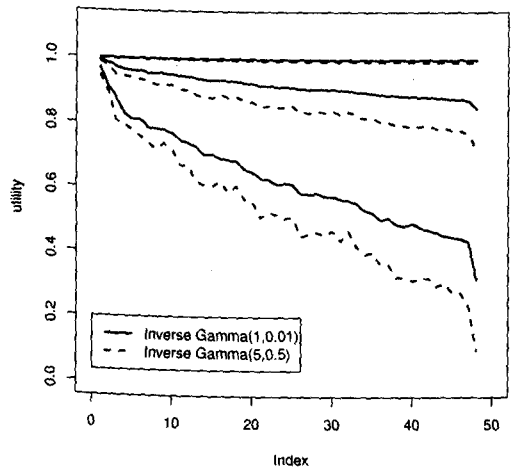
The mean utilities and 95% posterior intervals are calculated for the 48 health states in the sample using the results from the three prior distributions. Figures 6.7.2, 6.7.2 and 6.7.2 present the mean and 95% posterior interval of the 48 health states using the Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) prior distributions for the variance of the random effects. The mean utilities are similar using Inverse-Gamma(10,0.01) and Inverse-Gamma(5,0.5) prior distributions. The mean utilities using an Inverse-Gamma(1,0.01) prior distribution are larger. The upper posterior limits are all very close to 1. The posterior interval for an Inverse-Gamma(5,0.5) prior has the largest range.



Mean utility and posterior interval of 48 health states using Inverse-Gamma(10,0.01) and Inverse-Gamma(1,0.01) priors for variance of random effects] for iteration 70001 to 120000



Mean utility and posterior interval of 48 health states using Inverse-Gamma(10,0.01) and Inverse-Gamma(5,0.5) priors for variance of random effects using iteration 70001 to 120000



Mean utility and posterior interval of 48 health states using Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5) priors for variance of random effects for iteration 70001 to 120000

6.8 Conclusion

This chapter considers both maximum likelihood estimation and Bayesian inference for a random effects model using the AQL-5D data. Several starting values are considered for the maximum likelihood estimation. The maximum likelihood results depend on the starting values of the random effects, $\boldsymbol{\alpha}$. Small initial values of $\boldsymbol{\alpha}$ result in small estimates of the mean and variance of $\boldsymbol{\alpha}$ and larger estimates of $\boldsymbol{\theta}$ and σ. Larger initial values of $\boldsymbol{\alpha}$ result in a larger mean and variance of $\boldsymbol{\alpha}$ and relatively smaller estimates of $\boldsymbol{\theta}$ and σ. This suggests that $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are not identifiable and several combinations of values for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ produce the same likelihood value.

Using a Bayesian approach, a Gamma(1,10) prior distribution is assumed for the elements of $\boldsymbol{\theta}$ and σ, and a Log-Normal$(0,\tau)$ prior distribution is assumed for the elements of $\boldsymbol{\alpha}$. Three prior distributions are investigated for τ: Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5). An MCMC is run using each of the three prior distributions for two sets of starting values. The first set of starting values has a mean random effect close to 1, a small random effect variance and relatively large values of $\boldsymbol{\theta}$ and σ. The second set of starting values have a larger mean and variance for the random effects and small values for $\boldsymbol{\theta}$. The choice of prior has some influence on the posterior distribution. However, the starting values for $\boldsymbol{\alpha}$ has a substantial effect on the posterior distribution for τ. The larger the variance of the starting values for $\boldsymbol{\alpha}$ the further away from zero the distribution of τ is located. This suggests that, as in the maximum likelihood results, the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are not identifiable and several combinations of values for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ give the same likelihood value.

The random effects model is therefore not appropriate for the AQL-5D discrete choice data when fitting by either maximum likelihood estimation or Bayesian inference.

Chapter 7

Conclusion

The National Institute for Health and Clinical Excellence (NICE) is responsible for making recommendations about which drugs and treatments should be available on the NHS. As part of the decision making process, an economic evaluation is performed, which is measured in cost per QALY gained. The number of QALYs gained is calculated using the number of life years gained and a measure of preference, called a utility, for health states experienced during the years gained.

This thesis considers Bayesian inference to estimate utilities using discrete choice data. The health states are defined using the AQL-5D classification system which is an asthma specific system and defines 3125 health states. The AQL-5D data analysed in this thesis consists of four sets of eight pairwise comparisons. A total of 168 people valued one set of comparisons. Several models were fitted to the data in order to estimate the utility of any health state defined by the classification system. These models were either standard fixed effect models or multiplicative random effects models.

In the standard model three prior distributions were considered for the parameters. When using either a Gamma(1,10) prior or a Uniform(0,1) prior distribution

the mean and posterior intervals for each parameter are similar using both models. However, the use of a Gamma(5,5) prior distribution changes these results. Therefore when preference is given to smaller values in the prior distribution the results do not change from when all parameter values are equally likely. Results change when preference is given to larger parameter values in the prior and smaller values are less likely. This suggests that if a prior does not favour larger values the posterior distribution is robust to the prior distribution. In these models the same prior distribution is assumed for each parameter. However, as the amount each parameter decreases is expected to vary between the parameters, it might be useful to investigate the use of different priors for each parameter.

Bayes factors are used to compare the logit and probit models for each prior and to compare the logit and probit with the benchmark model. The Bayes factors show that both the logit model and probit model are suitable for the AQL-5D discrete choice data when compared with the benchmark model. The inclusion of an unusual observation in the comparison between death and the health state $(5, 2, 1, 1, 1)$ resulted in an unexpected large Bayes factor in favour of the probit model when compared with the logit model. When the observation is removed the Bayes factors were as expected, showing no difference between the two models. This problem showed that the evidence a single Bayes factor presents should only be considered when used in comparison with other Bayes factors.

Bayes factors are also used to compare the logit models with five reduced models, where one five attributes in the classification system is removed from the data in each model. Each attribute in the AQL-5D classification system helps explain the data and is important in defining utilities. The attribute activity limitations has been shown to have the greatest impact on the utility.

In the random effects model a Gamma(1,10) was assumed for the parameters

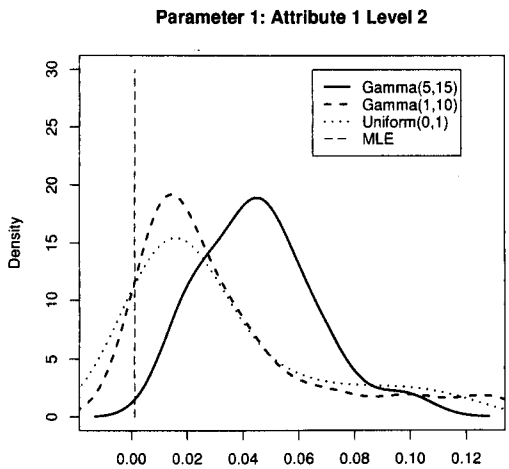
representing a decrease in utility, $\boldsymbol{\theta}$ and the scale parameter, σ. The random effects are assumed to have a lognormal distribution with a fixed mean equal to 1. Three prior distributions were considered for the variance of the random effects: Inverse-Gamma(10,0.01), Inverse-Gamma(1,0.01) and Inverse-Gamma(5,0.5). Two MCMC simulations were run with different starting values. The first set of starting values has a mean random effect close to 1, a small random effect variance and relatively large values of $\boldsymbol{\theta}$ and σ. The second set of starting values have a larger mean and variance for the random effects and small values for $\boldsymbol{\theta}$. The choice of prior has some influence on the posterior distribution. However, the starting values for the random effects, $\boldsymbol{\alpha}$, has a substantial effect on the posterior distribution for the variance τ. The larger the variance of the starting values for $\boldsymbol{\alpha}$ the further away from zero the distribution of τ is located. This suggests that the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are not identifiable and several combinations of values for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ give the same likelihood value.

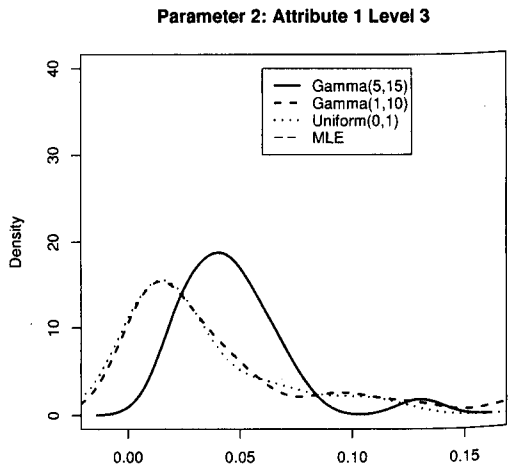
This thesis shows that the standard model is an appropriate model to use for the AQL-5D discrete choice data set. An improvement to the model is to consider different combinations of priors for the parameters. Bayes factors are useful to compare the standard logit and probit models when used in comparison with a benchmark model, and also to investigate the importance of each attribute in describing the utility. The multiplicative random effects model cannot be used for the AQL-5D data as the parameters are not identifiable and therefore the posterior distributions are influenced by the starting values of the random effects model. Some possible improvements could be considered. In the current model the mean of the random effects is assumed constant. An alternative model could allow this mean value to vary and a suitable prior distribution could be assumed. It is also a possibility that this method of assigning a random effect is never suitable for the logit model. The mixed logit model could be fitted which is defined in section 3.54. This model is an extension of the logit model but whereas in the

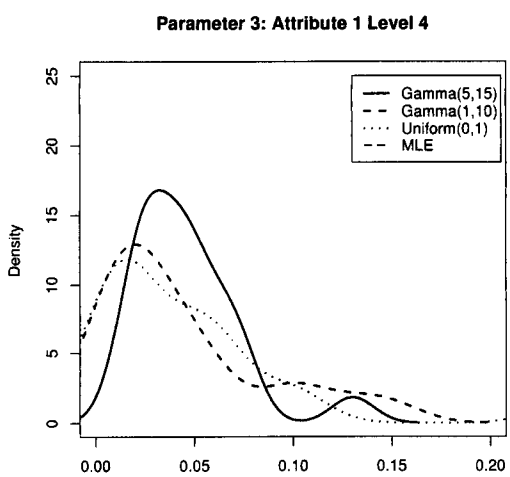
logit model the parameters are assumed to be constant, in the mixed logit the vector of parameters, $\boldsymbol{\theta}_i$ is assumed to have a particular distribution. The relationship between the mixed logit and the multiplicative random effects model is described in section 6.3.1. Priors could therefore be assumed for the parameters of the distribution assumed in the mixed logit model and the results compared with the multiplicative random effects model. In addition to these alternative, as the data or classification might not be suitable for a random effects model, the use of this model could be applied to data collected using a different classification system.

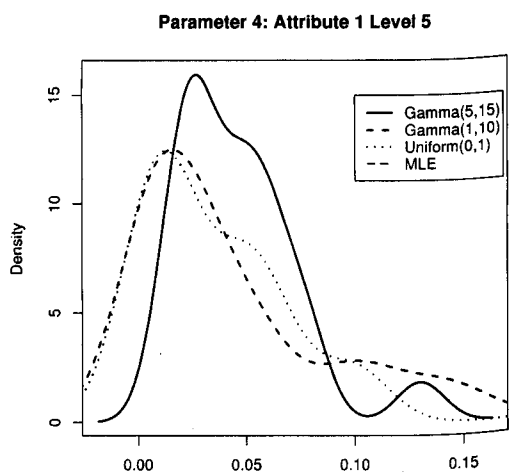
Appendix

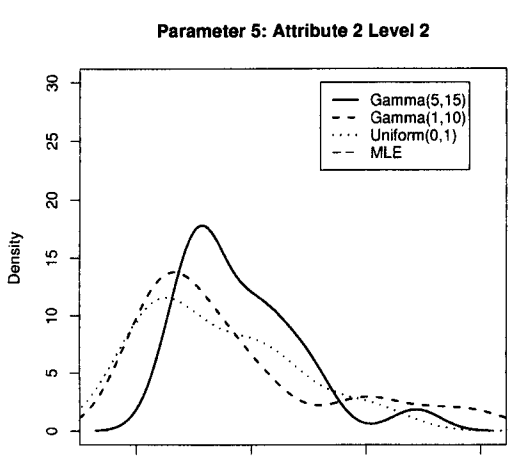
7.1 Logit Model Posterior Distributions for all Prior distributions

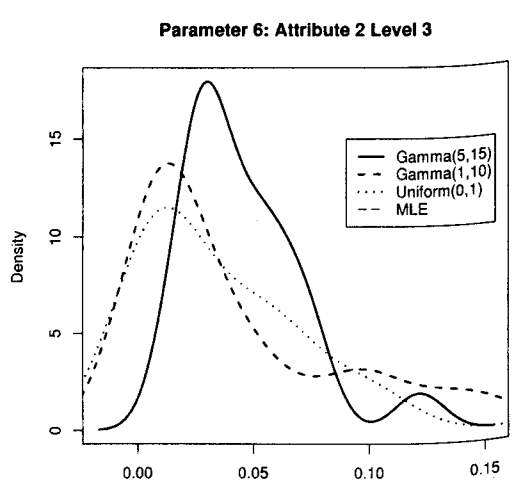




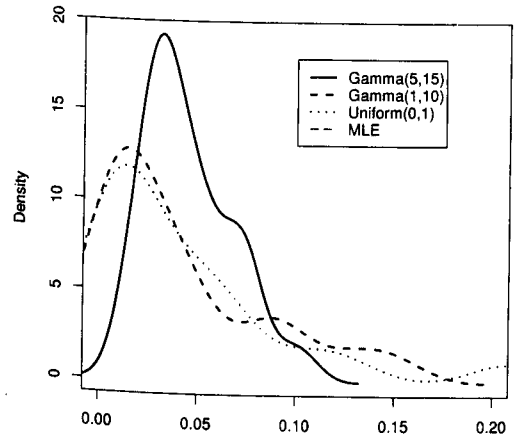




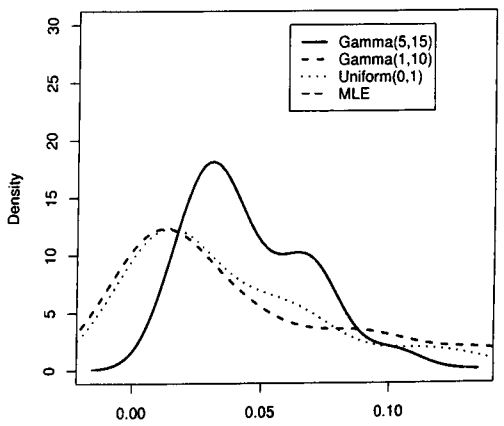




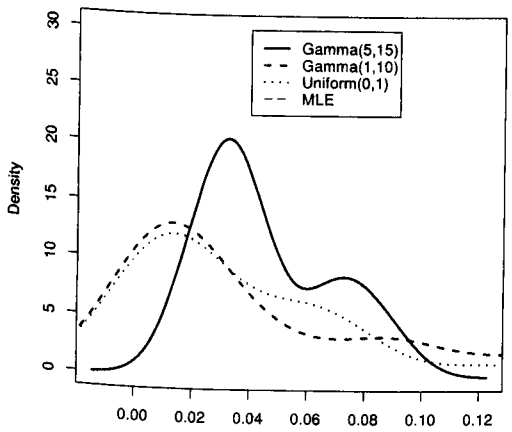
Parameter 7: Attribute 2 Level 4



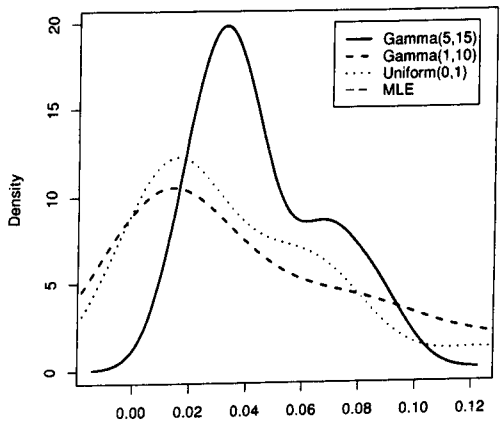
Parameter 8: Attribute 2 Level 5



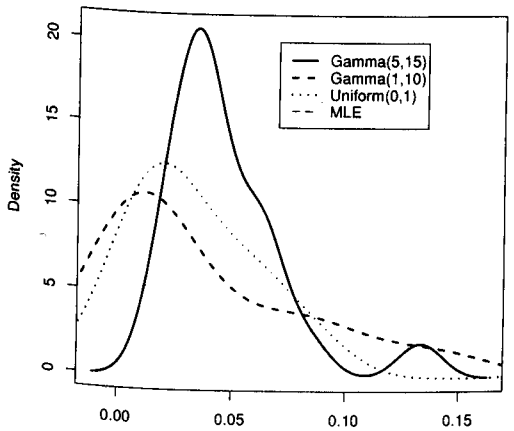
Parameter 9: Attribute 3 Level 2



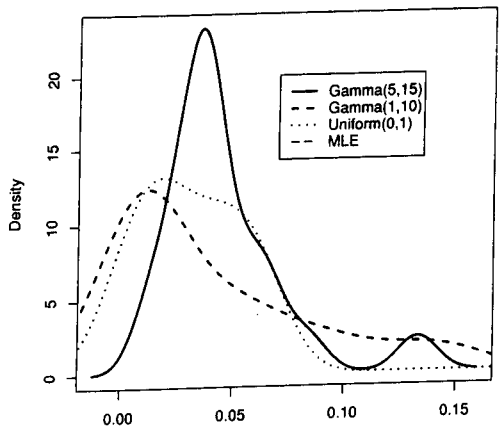
Parameter 10: Attribute 3 Level 3



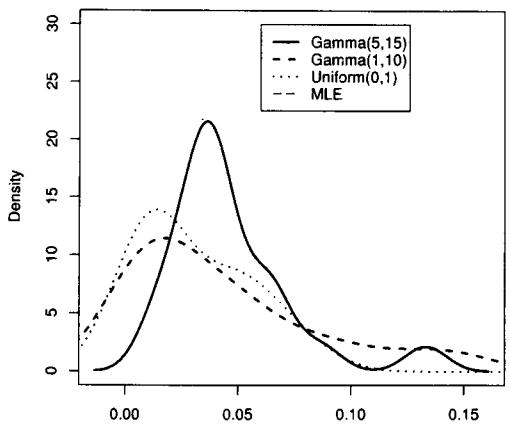
Parameter 11: Attribute 3 Level 4



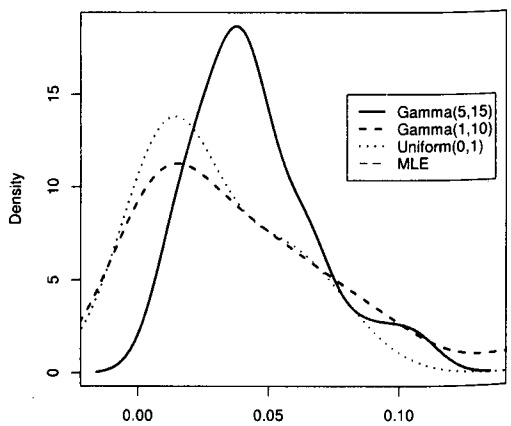
Parameter 12: Attribute 3 Level 5



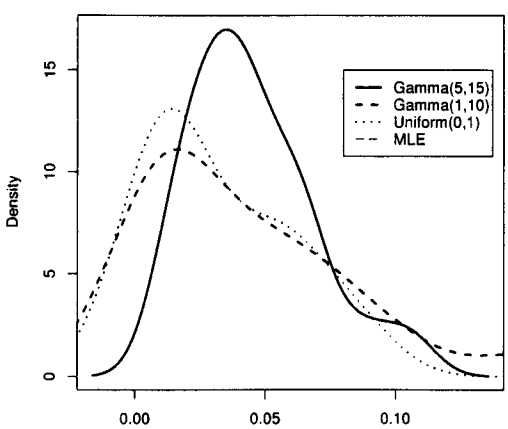
Parameter 13: Attribute 4 Level 2



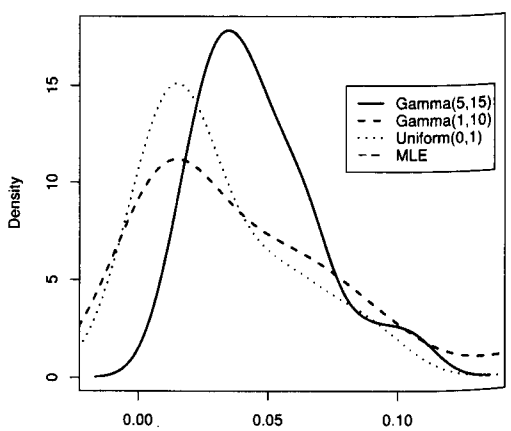
Parameter 14: Attribute 4 Level 3



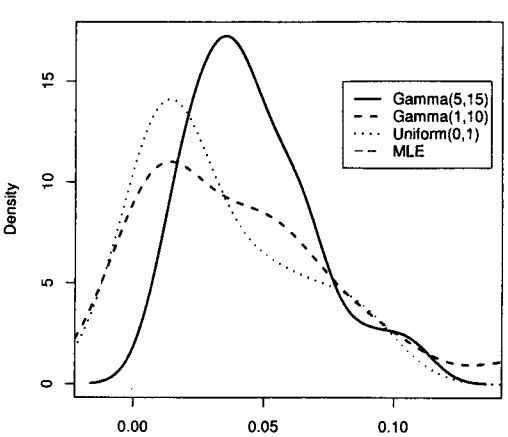
Parameter 15: Attribute 4 Level 4



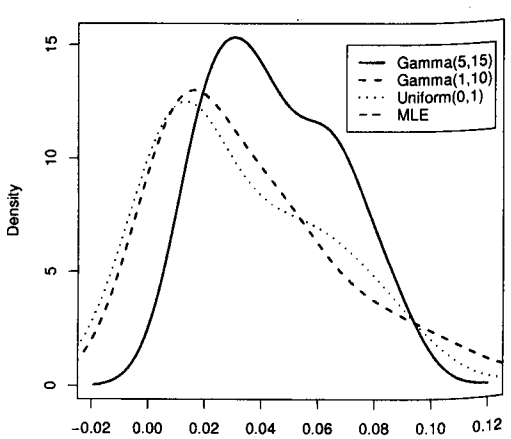
Parameter 16: Attribute 4 Level 5

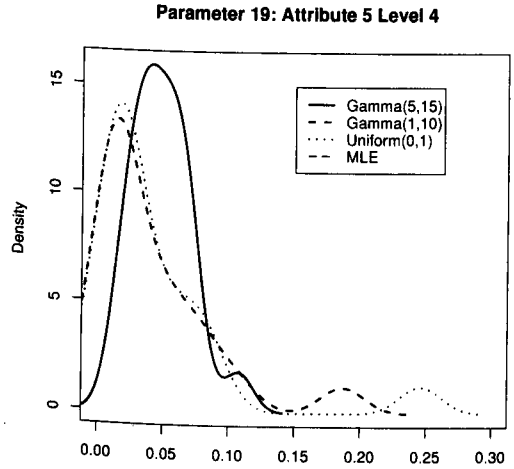


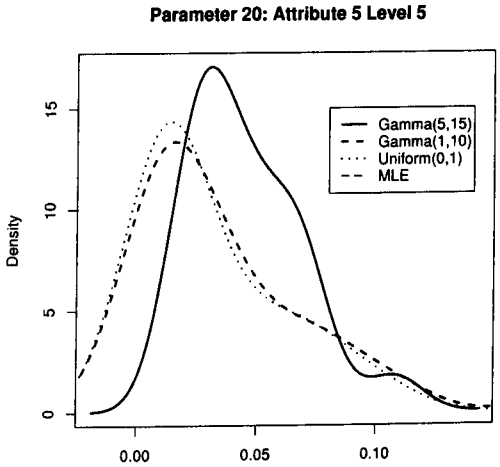
Parameter 17: Attribute 5 Level 2

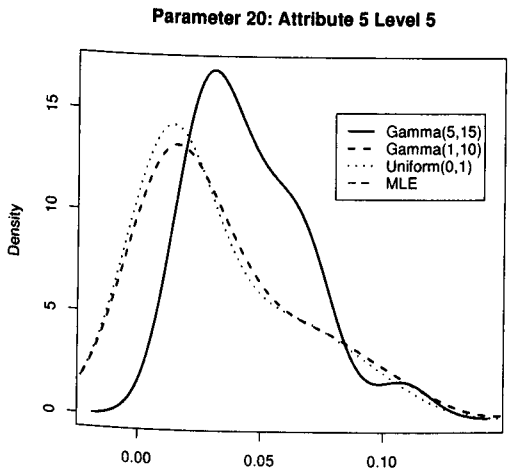


Parameter 18: Attribute 5 Level 3

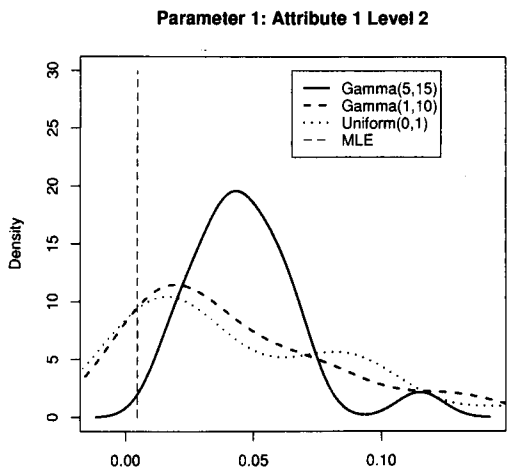


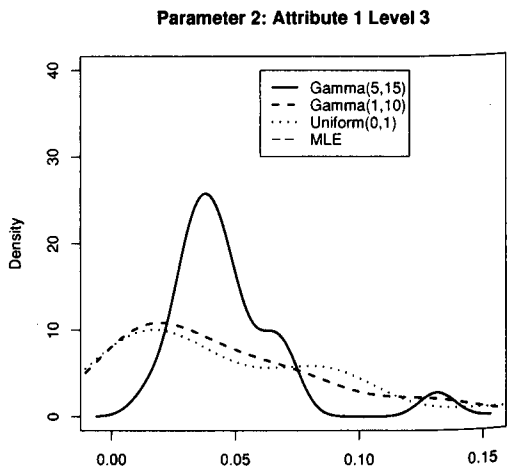

Parameter 19: Attribute 5 Level 4
Density
0
5
10
15
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE

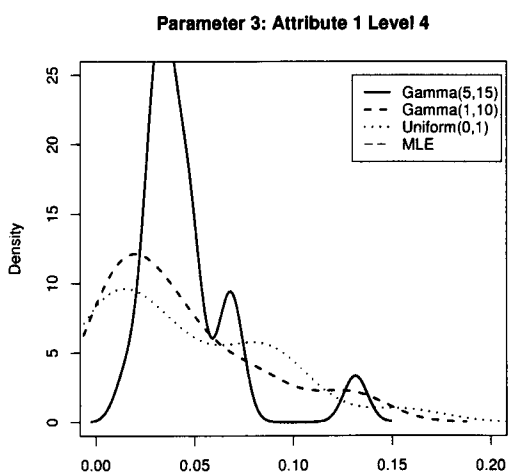

Parameter 20: Attribute 5 Level 5
Density
0
5
10
15
0.00
0.05
0.10
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE

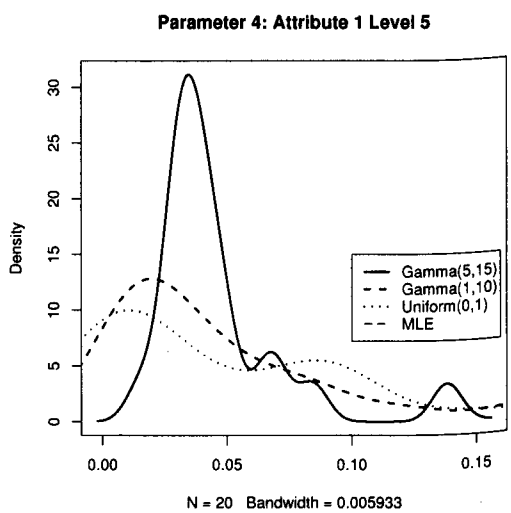

Parameter 20: Attribute 5 Level 5
Density
0
5
10
15
0.00
0.05
0.10
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE

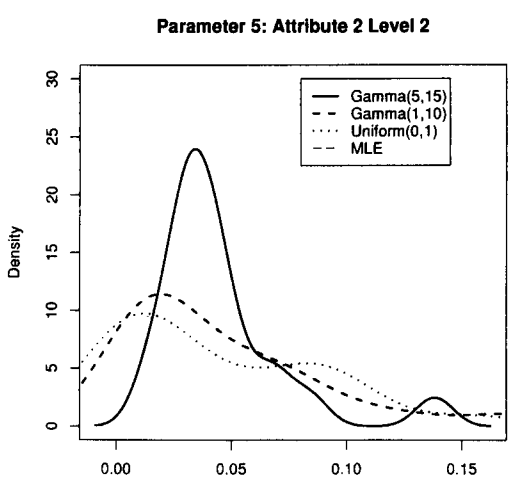
7.2 Probit Model Posterior Distributions for all Prior distributions

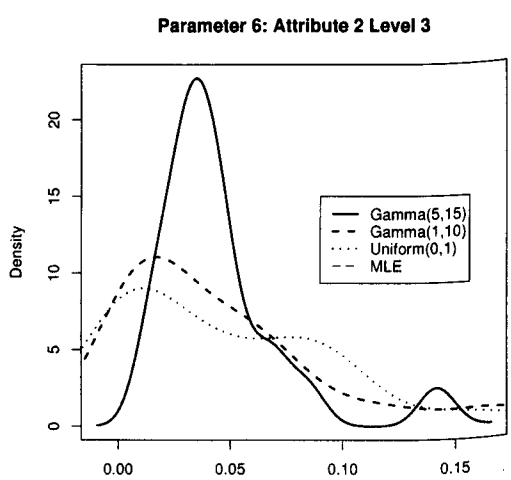




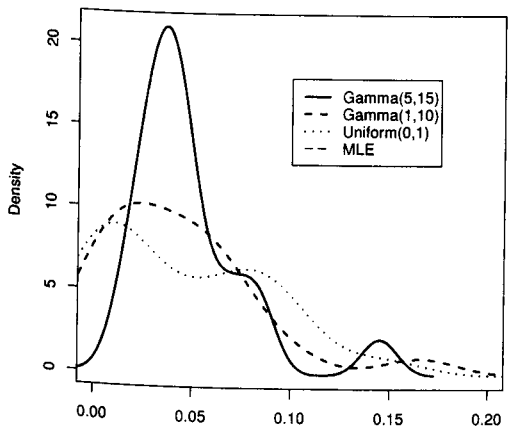




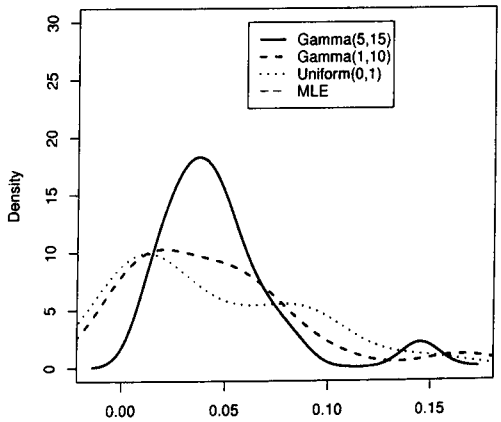




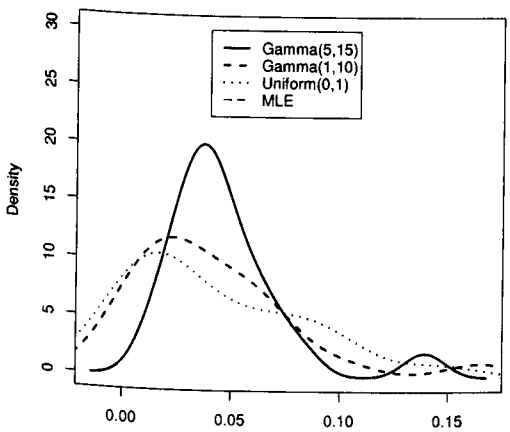
Parameter 7: Attribute 2 Level 4



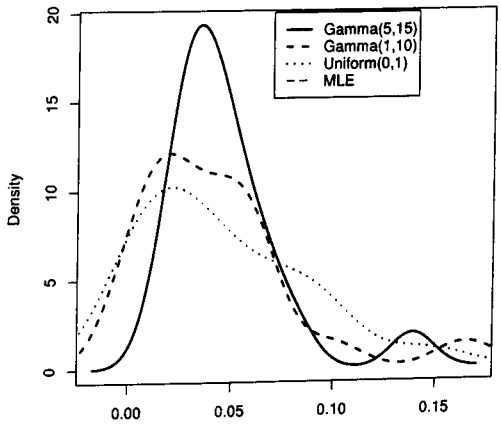
Parameter 8: Attribute 2 Level 5



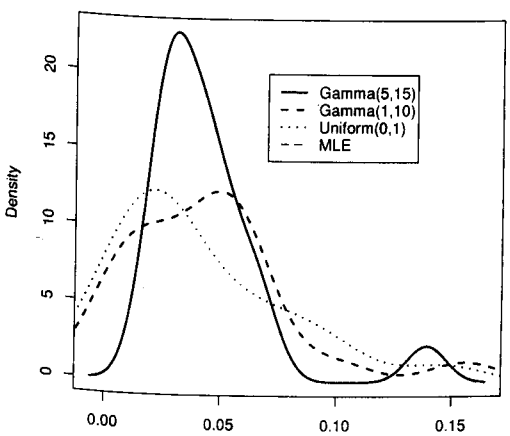
Parameter 9: Attribute 3 Level 2



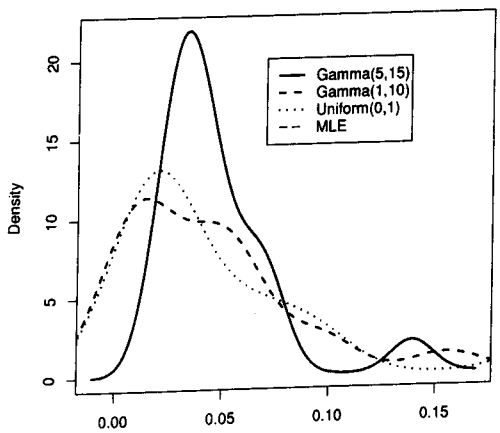
Parameter 10: Attribute 3 Level 3



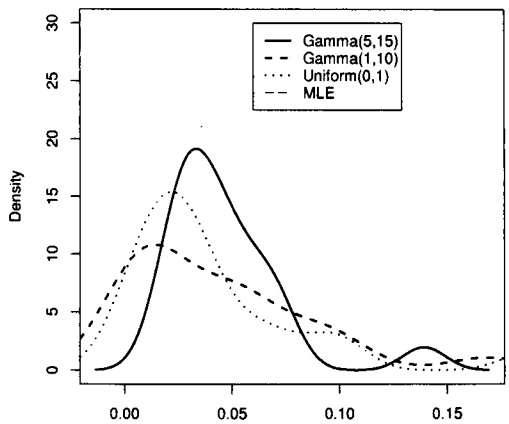
Parameter 11: Attribute 3 Level 4



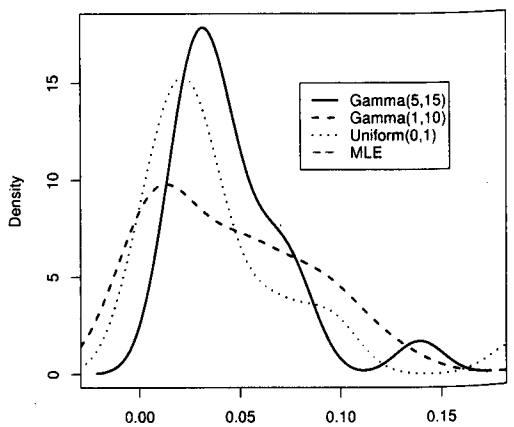
Parameter 12: Attribute 3 Level 5



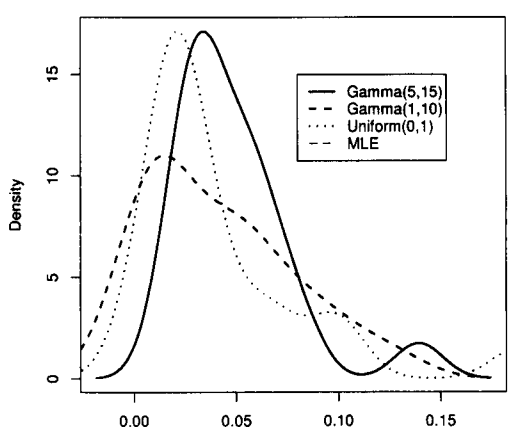
Parameter 13: Attribute 4 Level 2



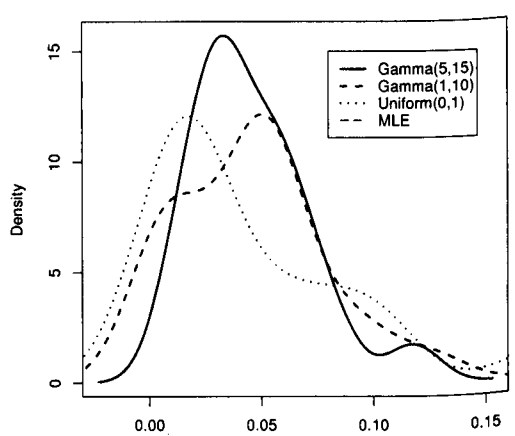
Parameter 14: Attribute 4 Level 3



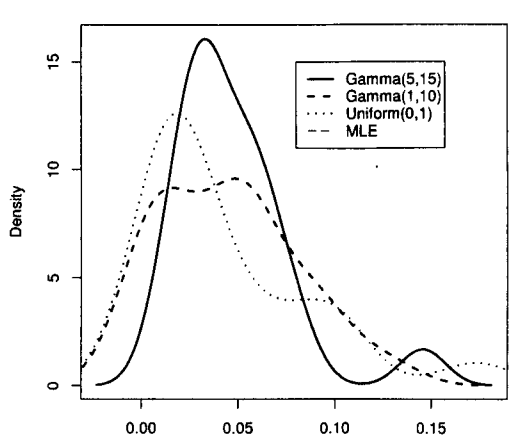
Parameter 15: Attribute 4 Level 4



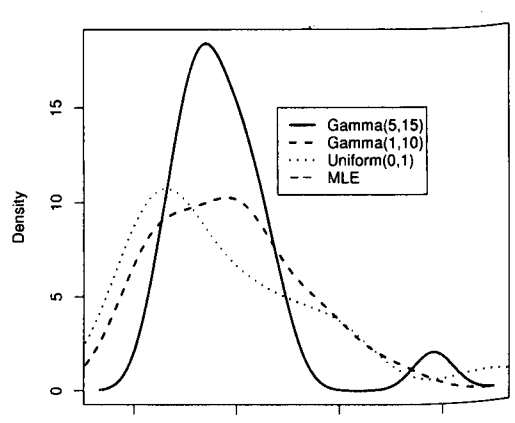
Parameter 16: Attribute 4 Level 5

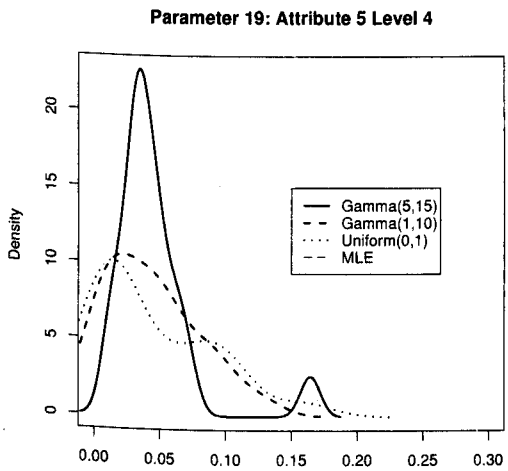


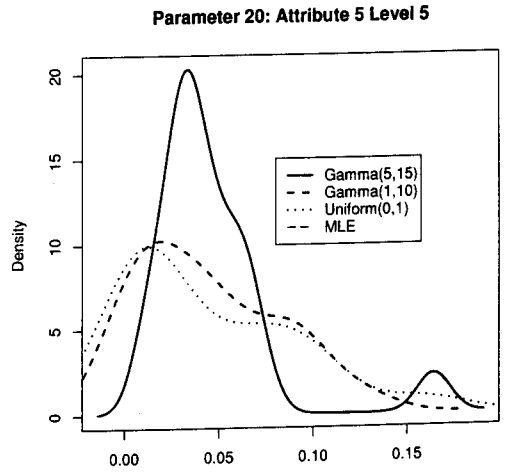
Parameter 17: Attribute 5 Level 2

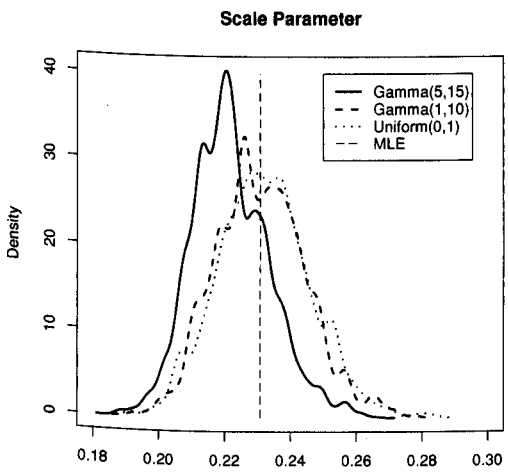


Parameter 18: Attribute 5 Level 3




Parameter 19: Attribute 5 Level 4
Density
0
5
10
15
20
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE


Parameter 20: Attribute 5 Level 5
Density
0
5
10
15
20
0.00
0.05
0.10
0.15
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE


Scale Parameter
Density
0
10
20
30
40
0.18
0.20
0.22
0.24
0.26
0.28
0.30
Gamma(5,15)
Gamma(1,10)
Uniform(0,1)
MLE

References

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data., *Journal of the American Statistical Association*, **88**: 669–678.

Barton, G., Bankart, J., Davis, A. and Summerfield, Q. (2004). Comparing utility scores before and after hearing-aid provision, *Applied Health Economics and Health Policy*, **3**: 103–105.

Basu, A., Dale, W., Elstein, A. and Meltzer, D. (2009). A linear index for predicting joint health state utilities from single health state utilities, *Health Economics*, **18**: 403–419.

Black, N., Griffiths, J. and Pope, C. (1996). Development of a symptom severity index and a symptom impact index for stress incontinence in women, *Neurourology and Urodynamics*, **15**: 630–640.

Brazier, J., Jones, N. and Kind, P. (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire, *Quality of life research*, **2**: 169–180.

Brazier, J., Ratcliffe, J., Salomon, J. and Tsuchiya, A. (2007). *Measuring and valuing health benefits for economic evaluation*, Oxford University Press.

Brazier, J., Roberts, J. and Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36, *Journal of Health Economics*, **21**: 271–292.

Brazier, J. and Tsuchiya, A. (2010). Preference-based condition-specific measure of health: what happens to cross programme comparability?, *Health Economics*, **19**: 125–129.

Brazier, J., Yang, Y. and Tsuchiya, A. (2006). Using rank and discrete choice data to estimate health state utility values for QALYs: the case of the AQL-5D, *Health Economics study group meeting, University of York*.

Brazier, J., Yang, Y. and Tsuchiya, A. (2008). A review of studies mapping (or cross walking) from non-preference based measures of health to generic preference-based measures, *Health Economics and Decision Science Discussion Paper 02/08*.

Brooks, R. and Group, E. (1996). EuroQol: The current state of play, *Health Policy*, **37**: 53–72.

Brooks, R., Jendteg, S., Lindgren, B., Persson, U. and Bjork, S. (1991). EuroQol: health related quality of life measurements. reaults of the swedish questionnaire exercise, *Health Policy*, **18**: 37–48.

Burr, J., Kilonzo, M., Vale, L. and Ryan, M. (2007). Developing a Preference-Based Glaucoma Utility Index using a Discrete Choice Experiment, *Optometry and vision science*, **84**: 797–808.

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., Brazier, J. and O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra, *Health Economics*, **14**: 339–347.

Dale, W., Basu, A., Elstein, A. and Meltzer, D. (2008). Predicting utility ratings for joint health states from single health states in prostate cancer: empirical testing of alternative theories, *Medical Decision Making*, **28**: 102–112.

Dolan, P. (1997). Modelling valuations for Euroqol health states, *Medical Care*, **35**: 1095–1108.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996a). The time trade-off method: results from a general population study, *Health Economics*, **5**: 141–154.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996b). Valuing health states: A comparison of methods, *Journal of Health Economcis*, **15**: 209–231.

Dolan, P. and Roberts, J. (2002). Modelling valuations for Euroqol health states: An alternative model using differences in valuations, *Medical Care*, **40**: 442–446.

Drummond, M., Sculpher, M., Torrance, G., O'Brien, B. and Stoddart, G. (2005). *Methods for the Economic Evaluation of Health Care Programmes*, Oxford University Press.

Espallargues, M., Czoski-Murray, C., Bansback, N., Carlton, J., Lewis, G., Hughes, L., Brand, C. and Brazier, J. (2005). The impact of age related macular degeneration on health state utility values, *Investigative Ophthalmology and Visual Science*, **46**: 4016.

EuroQol (1990). EuroQol: A new facility for the measurment of health related quality of life, *Health Policy*, **16**: 199.

Evett, I. (1991). Implementing bayesian methods in forensic science, *paper presented at the Fourth Valencia International meeting on Bayesian Statistics*.

Farrar, S., Ryan, M., Ross, D. and Ludbrook, A. (2000). Using discrete choice modelling in priority setting: an application to clinical service developments, *Social Science and Medicine*, **50**: 63–75.

Feeney, D., Furlong, W., Boyce, M. and Torrence, G. (1995). Multi-attribute health status classification systems: Heath utilities index, *Pharmacoeconomics*, **7**: 490–502.

Feeney, D., Furlong, W., Torrence, G., Goldsmith, C., Zhu, Z., DePauw, S., Denton, M. and Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system, *Medical Care*, **40**: 113–128.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*, Chapman and Hall/CRC.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, **57**: 1317–1339.

Gold, M., Siegel, J., Russell, L. and Weinstein, M. (1996). *Cost-effectiveness in Health and Medicine*, Oxford University Press.

Halekoh, U., Jorgensen, E. and Jensen, M. (2004). Ranking in discrete choice experiments, *Biometry Research Unit. Technical report.*

Haywood, K., Garratt, A., Lall, R., Smith, J. and Lamb, S. (2008). EuroQol EQ-5D and condition specific outcome measures in women with urinary incontinence: reliability, validity and responsiveness, *Quality of Life Research*, **17**: 475–483.

Hoeymans, N., van Lindert, H. and Westert, G. (2005). The health status of the Dutch population as assessed by the EQ-6D, *Quality of life Research*, **14**: 655–663.

Holbrook, M. and Skilbeck, C. (1983). An activities index for use with stroke patients, *Age and Ageing*, **12**: 166–170.

Huber, J. and Zwerina, K. (1996). The importance of utility balance in efficient choice designs, *Journal of Marketing Research*, **33**: 307–317.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability, *Proceedings of the Cambridge Philosophical Society*, **31**: 203–222.

Jeffreys, H. (1961). *Theory of Probability (3rd ed.)*, Oxford University Press.

Johnson, V. and Albert, J. (1999). *Ordinal Data Modeling*, Springer, New York.

Kass, R. and Raftery, A. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**: 773–795.

Keeney, R. and Raiffa, H. (1976). *Decision Making with Multiple Objectives*, Wiley: New York.

Keeney, R. and Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press.

Kessels, R., Goos, P. and Vandebroek, M. (2004). Comparing algorithms and criteria for designing Bayesian conjoint choice experiments, *Department of Applied Economics, Katholieke Universiteit Leuven, Belgium.*

Kessels, R., Goos, P. and Vandebroek, M. (2006). Optimal two-level conjoint designs for large numbers of attributes, *Department of Decision Science and Information Management, Katholieke Universiteit Leuven, Belgium.*

Kharroubi, S., Brazier, J., Roberts, J. and O'Hagan, A. (2007). Modelling SF-6D health state preference data: a non parametric Bayesian method, *Journal of Health Economics*, **26**: 597–612.

Kharroubi, S., O'Hagan, A. and Brazier, J. (2005). Estimating utilities from individual health preference data: a nonparametric Bayesian method., *Applied Statistics*, **54**: 879–895.

Kuhfeld, W., Tobias, R. and Garratt, M. (1994). Efficient experimental design with marketing research applications, *Journal of Marketing Research*, **31**: 545–557.

Lancsar, E., Hall, J., King, M., Kenny, P., Louviere, J. and Fiebig, D. (2003). Using discrete choice experiments to investigate patient preferences for preventive asthma medication, *Paper for ESAM03 Conference*.

Lemp, J. D., Kockelman, K. M. and Damien, P. (2010). The continuous cross-nested logit model: Formulation and application for departure time choice, *Transportation Research Part B*.

Longworth, L., Buxton, M., Sculpher, M. and Smith, D. (2005). Estimating utility data from clinical indicators for patients with stable angina, *The European Journal of Health Economics*, **6**: 347–353.

Luce (1959). *Individual Choice Behaviour*, Dover Publications Inc.

McCabe, C., Brazier, J., Gilks, P., Tsuchiya, A., Roberts, J., O'Hagan, A. and Stevens, K. (2006). Using rank data to estimate health state utility models, *Journal of Health Economics*, **25**.

McCabe, C., Stevens, K., Roberts, J. and Brazier, J. (2005). Health state values for the HUI2 descriptive system: results from a uk survey, *Health Economics*, **14**: 231–244.

McDonough, C., Grove, M., Tosteson, T., Lurie, J., Hilibrand, A. and Tosteson, A. (2005). Comparison of EQ-5d, HUI, and SF-36 derived societal health state values among spine patient outcomes research trial (SPORT) participants, *Quality of life Research*, **14**: 1321–1332.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour, in *Frontiers in Econometrics*, pp. 105–142, New York Academic Press.

Melzack, R. (1987). The short form McGill pain questionnaire, *Pain*, **30**: 191–197.

NICE (2008). Guide to the methods of technology appraisal.

Nord, E. (1991). EuroQol: health related quality of life measurements. valuations of health states by general public in norway, *Health Policy*, **18**: 25–36.

O'Hagan, A. and Forster, J. (2004). *Bayesian Inference*, volume 2B of "Kendall's Advanced Theory of Statistics". Arnold, London.

Plante, D., Piccirillo, J. and Sofferman, R. (1987). Decision analysis of treatment options in pyriform sinus carcinoma, *Medical Decision Making*, **7**: 74–83.

Read, J., Quinn, R., Berrick, D., Fineberg, H. and Weinstein, M. (1984). Preferences for health outcomes: Comparison of assessment methods, *Medical Decision making*, **4**: 315–329.

Robert, C. (2001). *The Bayesian Choice (2nd ed.)*, Springer.

Ryan, M. (2004). Discrete choice experiments in health care, *BMJ*.

Ryan, M., Bate, A., Eastmond, C. J. and Ludbrook, A. (2001). Use of discrete chocie experiments to elicit preferences, *Quality in health care*, **10**: 55–60.

Ryan, M. and Farrar, S. (2000). Using conjoint analysis to elicit preferences for health care., *BMJ*, **320**.

Salomon, J. (2003). Reconsidering the use of rankings in the valuation in the valuation of health states: a model for estimating cardinal values from ordinal data., *Population Health Metrics*.

Salomon, J. and Murray, C. (2004). A multi-method approach to measuring health state valuations., *Health Economics*, **13**: 281–290.

Spencer, A. (2003). The tto method and precedural invariance, *Health Economics*, **14**: 231–244.

Stavem, K., Froland, S. and Hellum, K. (2005). Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS, *Quality of Life Research*, **14**: 971–980.

Stewart, S., Lenert, L., Bharnagar, V. and Kaplan, R. (2005). Utilities for prostate cancer health states in men aged 60 and older, *Medical Care*, **43**: 347–355.

Sutherland, H., Llewellyn-Thomas, H., Boyd, N. and Till, J. (1982). Attitudes towards quality of survival: The concept of maximal endurable time, *Medical Decision Making*, **2**: 299–309.

Torrence, G. (1976). Social preferences for health states: an empirical evaluation ofthree measurement techniques, *Socio-Economic Planning Sciences*, **10**: 129–136.

Torrence, G., Thomas, W. and Sackett, D. (1972). A utility maximization model for evaluation of health care programs, *Health Services Research*, **7**: 118–133.

Train, K. (2003). *Discrete choice methods with simulation*, Cambidge University Press.

Tsuchiya, A., Brazier, J., McColl, E. and Parkin, D. (2002). Deriving preference-based single indices from non-preference based condition-specific instruments: Converting aqlq into eq5d indices, *Health Economics and Decision Science Discussion Paper 01/02*.

Viney, R., Savage, E. and Louviere, J. (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care, *Health Economics*, **14**: 349–362.

Wagner, T., Patrick, D., Bavendam, T., Martin, M. and Buesching, D. (1996). Quality of life of persons with urinary incontinence: Development of a new measure, *Urology*, **47**: 67–71.

Walters, S., Morrell, C. and Dixon, S. (1999). Measuring health-related quality of life in patients with venous leg ulcers, *Quality of life Research*, **8**: 327–336.

Yang, Y., Tsuchiya, A., Brazier, J. and Young, T. (2006). Estimating a preference-based index from the asthma quality of life questionnaire (AQLQ)., *Health Economists study group meeting*.