Mathematical Aggregation of Probabilistic Expert Judgements

Taposhri Ganguly

A thesis submitted in partial fulfillment for the

degree of Doctor of Philosophy

in the

Department of Management Science

University of Strathclyde

Supervisors: Prof. John Quigley and Prof. Lesley Walls

April 2017

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Taposhri Ganguly

Date: April 6, 2017

Abstract

Mathematical aggregation of the probabilistic expert's judgements in a structured expert judgement analysis is said to be relevant and critical. Due to absence of data, the judgements of the experts are used to perform forecasting and risk analysis. However, there is a gap in the literature where there may exist correlation in such judgements. This research is concerned with the situation where multiple experts are providing their numerical probability assessment for multiple quantities of interest. For each quantity of interest, there is a need of linear optimal weight on the basis of the experts' judgement. This optimality is achieved by minimising the mean squared error (MSE) between the unbiased judgements provided by the experts for a quantity of interest, whose true value is unknown. Further, it has been assumed that the judgements of the experts is dependent on the sets of multiple quantities of interests, while, their errors presented in the judgements are correlated. This thesis presents two novel mathematical methods towards aggregating expert's judgement through linear pooling. The first method is based on the empirical Bayes parametric formulation, and the second method is non-parametric. Both the chosen methods are compared using a stimulation study. This is to examine the performance of a given dependency structure which is further illustrated using a case study. In this context, a highly positive correlated

expert gets the least weight when compared to an independent or negatively correlated experts. As stated in literature and reaffirmed through the simulation studies in this thesis, asymptotically the non-parametric approach has a slower error rate convergence, where the error is defined in terms of the MSE in comparison to the parametric empirical Bayes method. Based on the simulation study and the case study results, it is found that the empirical Bayes method outperforms the non-parametric method.

Acknowledgements

This thesis would not have been possible without the wise, enthusiastic and endless academic support of my first supervisor, Professor John Quigley. I thank him for his trust in me to carry out this research and keeping me grounded while encouraging me to explore my intellectual quest. My supervision meetings with him have always left me excited, interested and super enthusiastic about research. I would like to thank my second supervisor, Professor Lesley Walls, for introducing me to the criticalities of the life of a researcher; as a doctoral student and beyond. Her invaluable feedback and incredible academic support has been crucial for this thesis. I would also like to thank her for sharing her vast experience and knowledge in the field of risk and reliability with me throughout my study. I am truly indebted to both my supervisors, Professors John Quigley and Lesley Walls, for finding time to discuss my work even beyond the formal working hours.

I would like to thank Ms Alison Kerr, my departmental administrator, for sorting my life at the department and for chatting to me most optimistically about my progress through the years.

I would like to thank all my former and current PhD colleagues at the department for keeping and maintaining high standards of research; especially Ms Kanogkan Leerojanaprapa, Mr

Alistair Scotland and Mr Mouhamad Shaker Ali Agha; for their wonderful company, all the useful chats surrounding research philosophy, methods, statistics; for restaurant hunts, and for spending nights at the department working over some fish and chips.

I would like to thank the department of Management Science, University of Strathclyde for funding my research and for all the help, support and advice that I was provided throughout my time at the department and the university.

I am grateful to Professor Roger Cooke for taking a keen interest and discussing my thesis with me, raising my interest in modelling conditional dependencies and sharing his invaluable knowledge in the field of expert judgement. My discussions with him have left me excited about research.

And lastly, I owe everything to my parents. I would not be here if it had not been for their unwavering faith, kindness, support, love and trust.

Contents

List of Figures

List of Tables

Notation

z represents the number of experts

n represents the total number of questions

i, j represent the experts where i and $j \in 1, 2, ..., z$

k represents the question number; it acts as a counter for the number of questions, therefore $k \in 1, 2, ..., n$

X_{ki} is the random variable representing expert i's assessment on question k

x_{ki} is the realisation of X_{ki}

$\mu_{ki} = E\left(X_{ki}\right)$ is the expectation of expert i's assessment of question k

$\mu_i = E\left(X_{ki}|\mu_{ki}\right)$ is the conditional expectation of expert i's assessment of question k and is the true value for expert i's assessment $\forall\ k$

$\sigma_{ki} = \sqrt{E\left(X_{ki} - \mu_{ki}\right)^2}$ is the standard deviation of expert i's assessment on question k

$c_{ijk} = Cov\left(X_{ki}, X_{kj}\right)$ is the covariance between expert i and expert j's assessment over k^{th} question; where $Cov\left(X_{ki}, X_{kj}\right) = E\left(\left(X_{ki} - \mu_{ki}\right)\left(X_{kj} - \mu_{kj}\right)\right)$.

C_k is the covariance matrix for k^{th} question where c_{ij} is the $(i, j)^{th}$ element $\forall\ k$

C represents the covariance matrix $\forall\ k$

C^{-1} represents the inverse of the covariance matrix $\forall\ k$

$\rho_{ijk} = \dfrac{Cov\left(X_{ki}, X_{kj}\right)}{\sigma_{ki}\sigma_{kj}}$ is the correlation between expert i and expert j's assessment on question k

$\theta_k = E\left(\mu_{ki}\right)$ is the prior mean of expert i's assessment of question k

$\tau_k = Var\left(\mu_{ki}\right)$ is the prior variance of expert i's assessment of question k

Π is the correlation matrix whose elements are σ_{ki}, c_{ij}, μ_k and ρ_{ij}

w_i is the weight assigned to expert i

$\underline{\mathsf{w}} = (w_1, ..., w_z)$ is the vector of weights assigned to experts

λ is the Lagrange multiplier

$\hat{\rho_{ijk}}$ is the estimator of ρ_{ijk}

$\hat{\rho_{ij}} = \dfrac{\sum\limits_{k=1}^{n} \hat{\rho_{ijk}}}{n}$

$\hat{w}_i$ is the estimator of w_i

$\hat{\theta}_k$ is the estimator of θ_k

$\hat{\tau_k}$ is the estimator of τ_k

$\hat{\mu_{ki}}$ is the estimator of μ_{ki}

$\hat{\mu}$ is the sample average

U is the functional space

L_c represents the copula function

List of acronyms

$E1$: Expert 1

$E2$: Expert 2

$E3$: Expert 3

$E4$: Expert 4

L : Lagrange constraint

MVN : Multivariate normal Distribution

BVN : Bivariate normal Distribution

MSE : Mean square error or mean squared error

$OL/B/FE$: Conditions on capacity, tensioning and grout condition: Original Level, Bend and Fully effective

$OL/B/PE$: Conditions on capacity, tensioning and grout condition: Original Level, Bend and Partially effective

$OL/B/NE$: Conditions on capacity, tensioning and grout condition: Original Level, Bend and Not effective

FE : Fully Elastic

JUD : Juddering

NM : No Movement

TF : Total Failure

F: Failure

$FW/E/EL$: Failed with evidence elasticity

$FW/NE/EL$: Failed with no evidence elasticity

$RC/AL/B/FE$:Conditions on capacity, bend and grout: Reduced Capacity Acceptable Level,Bend, Fully effective

$RC/AL/B/PE$: Conditions on capacity, bend and grout: Reduced Capacity Acceptable Level, Bend, Partially Effective

$RC/AL/B/NE$: Conditions on capacity, bend and grout: Reduced Capacit Acceptable Level,Bend, Not effective

$RC/AL/NBD/FE$: Conditions on capacity, bend and grout: Reduced Capacity Acceptable Level, No bend, Fully effective

$RC/AL/NB/PE$: Conditions on capacity, bend and grout: Reduced Capacity Acceptable Level, No bend, Partially Effective

$RC/AL/NB/NE$: Conditions on capacity, bend and grout: Reduced Capacity Acceptable Level, No bend, Not effective

$RC/UL/B/FE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, Bend, Fully effective

$RC/UL/B/PE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, Bend, Partially Effective

$RC/UL/B/NE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, Bend, Not effective

$RC/UL/NB/FE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, No bend, Fully effective

$RC/UL/NB/PE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, No bend, Partially Effective

$RC/UL/NB/NE$: Conditions on capacity, bend and grout: Reduced Capacity to Unacceptable level, No bend, Not effective

$F/B/FE$: Failure,Bend, Fully effective

$F/B/PE$: Failure, Bend, Partially Effective

$F/B/NE$: Failure,Bend, Not effective

$F/NB/FE$: Failure, No bend, Fully effective

$F/NB/PE$: Failure, No bend, Partially Effective

$F/NB/NE$: Failure, No bend, Not effective

$Npar$: Non-parametric

EB : Empirical Bayes

$MCMC$: Markov Chain Monte Carlo simulation

Chapter 1

Introduction

1.1 Research motivation

Judgements and decision-making are widely discussed and implemented across many disciplines such as social psychology, behavioural economics and management science, to tag a few. Judgements are often defined as the processes that lead to decision making. Howard, (1998) stated that the main reason for such a wide discussion across the disciplines and within psychology is due to the decision-making that lies in every corner of the human thought process. O'Hagan et al (2006) added that another such reason is the ability to make the decisions which is often attributed to the cognitive processes within the human mind.

Decision making, especially under uncertainty paves the way for practical and philosophical thought governing processes. Further, it leads to the development of decision analysis. The situations that require decision making vary and so does the complexity of it. Decisions are often based on human judgements and beliefs and sometimes based on experimental evidence that the individual has been exposed to. Kahneman and Tversky (1982) provided the reasons for the human mind's ability to make judgements under uncertainty. Their list comprised of reliance on judgemental heuristics and prevalence of biases that the authors felt were responsible for shaping the mind. Thus, as per the assumptions of Howard (1998), it can be said that the discipline of decision analysis follows a course of action or a course of thought process, which is governed by subjective beliefs. This fundamental aspect of the discipline finds its roots embedded in the statistical decision theory. The statistical decision theory also provides a course of action that leads this thought process to form a decision under uncertainty or according to the demands of the problem under consideration (Howard, 1998). At the heart of decision analysis, there lies a process that is governed by information and refers to the models, relationships or the probability assignments. This might be important in characterising the links between outcomes and decisions. These models could be complex and dynamic or even relatively simple in many cases, and the remaining uncertainty would be characterised by probability assignments (O'Hagan et. al, 2006). The very task of decision making is not often easy for the human nature, considering the process of logical thought in applying a belief to the value of the system or a process (Hogarth, 1978, 1987, 2001; Girotto, et. al, 2001). Further, in an uncertain world, reasonable decisions may lead to non-desired outcomes and vice versa. As a result, the distinction between decision and outcomes becomes very crucial to perceive, and it forms an important criteria which is needed for decision analysis. It is quite sure that the modelled probability distributions follow a specific numerical construct and mathematical modelling assumptions while representing an approximation of the empirical rational system. This can further be used to understand and predict uncertain events (Cooke, 1991). However, Kahneman and Tversky (1982) opined that an individual is often coherent with his/her judgements. The authors further touched upon the subjectivity that governs the beliefs conveyed on a particular system or the probability values to a prescribed problem. Then, they concluded that the subjective probabilities play a crucial role in life because of its very nature. The people also tend to apply the correct Bayesian rules to the decision problems based on its intuitive nature, but often fail to appreciate the full statistical principle behind the rules. Although this inner coherency may or may not lead to a good outcome, but given the uncertainty, there is a crucial need to account for such coherency.

Surowiecki (2005) mentioned that one of the core ideas regarding decision making that was put forward by philosophers, including Aristotle, was the wisdom that lies in crowds as opposed to that of a single person. At a very basic level, it stresses upon the fact that the two heads are better than one, and if there are more number of heads, then, the better outcome of a decision is anticipated as noted by Lyon and Pacuit, (2013). The study of the wisdom of crowd spans across disciplines, with a wide range of literature across management science, computer science, social psychology, social choice theory and behavioural economics, to name a few. A major part of management science deals with the issues of decision-making, and this process often involves three key entities, namely the decision-makers, the analysts and the experts. More often, the analysts and the decision-makers refer to the individuals where a decision maker can play the role of an analyst based on the identified problem and vice versa. Budescu and Chen (2014) defined the 'wisdom of the crowd' as the ability of statistical aggregation based on the multiple opinions to outperform individuals, including experts, in various prediction and estimation tasks. Lyon and Pacuit (2013) revisited this definition and provided a brief insight into the aggregation of human judgements. They provided six core aspects; namely, the input, the output, the elicitation method, the aggregation, the recruitment, and the standards of evaluation. It may also be pointed out here that this collective wisdom is often used to solve a problem in management science and other disciplines, and its roots lie in the statistical decision theory as mentioned earlier.

There have been several statistical theories that have been formulated for decision making under risk and uncertainty. Assertions like these rely on the assumptions that there are no systematic errors in the assessments of the crowds. This is not true always. For an instance, in the field of experimental psychology, the empirical studies have presented the predictable errors made by the laypersons assessing the number of fatalities per annum due to the cause of death. On the other hand, the low frequency events are over-estimated whereas the higher ones are underestimated (Lichtenstein S, et. al, 1978). This highlights the need for a careful selection of experts who are knowledgeable about the problem under considerations. However, unfortunately, experts are not immune to such bias in their assessment, and this empathises the role of the elicitation process. Even with a well-constructed elicitation process, there can be a systematic error in judgement amongst the subsets of experts, derived from shared or similar experiences. This research is concerned with aggregating such judgement amongst experts where the systematic errors may be present amongst some.

Surowiecki (2005) presented an anecdote that highlighted the power of collective thinking over individual thinking with Sir Francis Galton's surprise on asking a lay person to guess the weight of an ox. At the conclusion of this activity, an individual could not correctly guess the weight of an ox but a crowd median guess was fairly accurate. The issue surrounding the collective thinking also highlighted the diversity that existed within the crowd, and the way the problem was perceived by each and every individual separately. This diversity in the thought process gives a wider range of insight into the domain knowledge that further helps to accurately solve a problem or as in this case helps to arrive at an accurate value for the weight of an ox, which matched its true butchered weight. Thus with the help of aggregation based on these human judgements, one can predict better or arrive at a better decision. When the 'crowd' is replaced from a general problem and appraisal of aggregation of human judgements is shifted to a specific domain knowledge, the definition of 'crowd' gets changed. It then maps onto a team of experts or a group of experts. These experts are defined as people who have strong domain knowledge and are regarded as specialists in the chosen area of work or the area where the problem exists (O'Hagan et. al, 2006). The judgements tend to be correlated with each other due to the shared domain knowledge and though the diversity is often desired, it does not always presents what it is exists. On the basis of these given dependencies, there remains a complete lack of knowledge regarding the true values of the problems which are desired to be solved. The challenges related to mathematics aggregate to provide a meaningful decision which later on becomes fairly challenging.

'Aggregation' refers to aggregating the subjective probability distributions from various experts and assessing the same scenario to obtain one subjective distribution. The challenge is to obtain a comprehensive and exhaustive database of information where the opinions are statistically independent. This is because the relevance and usefulness of information decrease with an increase in the correlations between the experts, as the information no longer remain statistically independent (Broomell and Budescu, 2009). In such scenarios, consulting multiple experts could be considered same as consulting one expert. This thesis is focussed on the development of mathematical models that aim to aggregate or combine these judgements accounting for the correlation in the error of these judgements. This aggregation provides the decision maker with one single probability distribution, which may be used to tackle decision problems. These models incorporate not only the positive dependencies between the judgements or assessments but also provide an aggregation framework when there is a potential disagreement within the group of experts.

1.2 Research gap

There is a gap in the literature between the aggregation of expert's judgement where judgements are correlated in errors. There is an evident gap in the literature also on the use of the mathematical approaches that have been proposed in this thesis. The data structure as expressed in Table 1.1 has not been aggregated based on the proposed mathematical methods, namely the empirical Bayes method and the non-parametric method using constrained optimisation with Lagrange multipliers in the literature. Further, the empirical Bayes method has not been explored in the literature towards aggregating expert judgements, accounting for the correlation in the judgement errors. A fairly interesting question then arises to explore why expert judgement aggregation lacks the potential use of aggregation using the empirical Bayes method for modelling statistical dependence [1] and also to provide mathematical models that would meaningfully combine these correlated judgements. The judgements have often been treated as the data for analysis within risk and reliability. Apart from that, often several semi-parametric and non-parametric mathematical approaches have been used towards aggregating these judgements, but none of the approaches has used the techniques of reducing the mean squared loss error. Therefore, the gap in the literature on the basis of the treatment of judgements accounts for the dependencies that exist between them and paves the way for this research.

1.3 Research aim

The aim of this research is to mathematically aggregate the probability judgements that the experts have provided based on the quantities of interest while accounting for the correlation in the error of these assessments. Table 1.1 shows the assumed structure of the problem that has been considered in this thesis. Each question measures some quantity of interest and each expert has provided his best guessed on that particular quantity of interest. It is assumed that each expert is unbiased therefore their best guess represents the true value of the underlying quantity of interest.

[1] Statistical dependence refers to the correlation in the judgement error throughout the thesis and dependence and correlation are used interchangeably throughout the text.

Table 1.1: Problem structure of the thesis in terms of experts' assessments of the questions

	E_1	E_2	...	E_i	E_z
Q_1	x_{11}	x_{12}	...	x_{1i}	...
...	...	...	...	...	...
Q_k	...	...	...	x_{ki}	...
Q_n	$x_{n1},$	...	...	...	x_{nz}

where x_{1i} is the realisation of expert i on question 1, as defined in the notation list. Given these realisations we can then compute the standard deviation of expert i's assessment on question k along with the expectation of expert i's assessment on question k. The formulas for computing the expectations and the standard deviations are listed under notation. The aggregation of these assessments is then achieved by two different mathematical methods. These methods have been adopted and presented in the thesis; one is non-parametric method towards the aggregation of dependent expert judgements and the other is an empirical Bayes method. The data considered throughout the thesis is the same structure as expressed in Table 1.1. The expert assessments are the realisations in the form of their best guesses and these are the point estimates. For the non-parametric method, a constrained optimisation technique using Lagrange multipliers has been proposed to minimise the mean squared error whereas, in the case of empirical Bayes method, it is assumed that these realisations are from a multivariate normal distribution with unknown parameters. There are further assumptions on the likelihood function and using the Bayes rule, the posterior distribution is then calculated. The non- parametric method is more appealing to be used in practice due to its ease of understanding and formulation than the empirical Bayes approach. These two different methods have been discussed in detail in chapters three and four and are compared and contrasted within the thesis in chapter five. Both of these methods have also been applied to a real expert's judgement data from an ageing bridge.

1.4 Research philosophy and methodology

In statistics, two main and contrasting philosophical paradigms have always existed namely, the frequentist and the Bayesian paradigm; however at the methodological level, these differences are often unexpressed (Bayyari and Berger, 2004). The frequentist school of thought runs on the belief that there exists a single truth. Besides that, the collected information forms a noisy sample of realisation of the true values using the process of repeated experimentations and sampling. The Bayesian school of thought, on the other hand, advocates the belief that there is no truth which can be assessed through the data collection (Ambaum, 2012).

The Bayesian belief is subjective and is conditioned on a random set of events which have their own subjective probabilities (Ambaum, 2012). It is evident in literature that human judgements have often been shelved as 'Bayesian', and the reason for this is attributed to the fact that judgements are beliefs, which are subjective by their very nature, and at many levels have an impact on the shared knowledge and experience in its collective form (Kahneman and Tversky , 1982; Hartmann and Hajek, 2010). At the outset, when an expert is asked to provide his beliefs, he provides his understanding of the subject in terms of his probabilities, which are not the part of any experiment; however, they could be based on his personal expertise and experience (Lindley, 2000). The debate based on the superiority of the frequentist or Bayesian with the others have received much attention in the literature. But, keeping the academic significance on mind, an objective related to Bayesian methodology is preferred over the individual paradigms of isolation (Bayyari and Berger, 2004; Efron, 2006). Several important statistical inferential theories are worthy of note (Cox and Hinkley, 1979); namely, the sufficiency principle, the likelihood principle, the invariance principle and conditionality principle. The likelihood principle is based on the notion that all the information from the sample is contained in the likelihood function thus constructed. The combination of the likelihood principle with the law of likelihood often produces the most probable values where the likelihood is maximised. This method of estimating parameters is known as the maximum likelihood method.

In this research, the available data is in the form of a secondary dataset [2] and are treated as data that belong to a certain class of probability distribution, which positions this thesis within a frequentist paradigm of statistics. The inferential theory used in this thesis follows the likelihood principle and the likelihood functions. This summarises all the information contained in the sample dataset. Though this thesis is built on a secondary dataset which is a collection of expert's judgement, but as these judgements are observed for repeated set of questions and the parameters remain constant throughout the repeatable process, the statistical philosophy adopted is in line with the frequentist paradigm. To reiterate, the critical positioning of this research is within the frequentist paradigm of statistics and the scope of this thesis is restricted to provide a better statistical inference based on the

[2] A dataset that is not collected for the purpose of this research and is not collected by the user of the dataset

repeated measures of these judgements provided by the experts through a repeatable sample of questions.

The empirical work in this research is based on an investigation of secondary data. that has resulted from the study of the reliability of the Forth Road Bridge conducted by Professors Quigley and Walls (2010). The use of secondary data in research has been speculated in literature; while there are several advantages of using secondary data. At the same time, there are some serious drawbacks where the data might not have been collected to answer similar research questions. However, as Ghauri (2004) argued that any research must start with secondary data, within this thesis as the research motivation is drawn from the data and the mathematical methods that are developed using the secondary dataset. But, later they are generalised (Ghauri, 2004). The outline of the research methodology adopted in this thesis is:

1. Modelling framework: Two mathematical models are proposed; namely, parametric model based on empirical Bayes formulation and non-parametric modelling framework for the aggregation of dependent judgements.

2. Model evaluation: The models developed are evaluated based on the statistical measures of mean squared errors and simulations.

3. Real world data application: The models thus developed, have been applied to the real world dataset of an ageing bridge.

1.5 Research contribution

The potential contributions to knowledge that this thesis aims to provide are enumerated as below:

1. To provide mathematical models that aggregate expert's judgement where the judgements are correlated in errors.

2. To demonstrate the impact of correlated errors on aggregation of these judgements not only in abstract examples but also in a real data of an ageing bridge. Thus, strengthening the impact of this research work in a real case of reliability and risk-analysis.

1.6 Outline of the thesis

This thesis is organised into seven chapters. Chapter two provides a literature review on aggregation of expert's judgements and highlights the gap in the literature. This chapter also provides a literature review of the proposed methods highlighting their strengths, weaknesses and usage. Chapter three presents the first mathematical method which is the non-parametric method towards the aggregation of correlated expert's judgements. This chapter also demonstrates the aggregation of judgements based on artificial examples. Chapter four presents the second mathematical method based on the empirical Bayes method for the aggregation of expert's judgements. This chapter demonstrates the use of empirical Bayes method towards aggregation based on the same examples from chapter three. Chapter five presents a simulation study comparing the approaches developed and proposed in chapter three and chapter four. Chapter six presents data from an ageing bridge where the mathematical methods developed are then illustrated. The last chapter that is chapter seven is a discussion chapter that argues the limitations of this thesis highlighting the key findings and future direction that this research could be potentially led to.

Chapter 2

Review of the expert judgement aggregation literature

2.1 Introduction

This chapter is divided into two main sections. Section 2.1 provides a literature review of the expert's judgement aggregation literature. The section is organised into various sections highlighting the behavioural and the mathematical models that exist in the literature based on the expert's judgement aggregation. Section 2.2 provides a literature review of the proposed models that are further developed for the aggregation of experts' judgement in this thesis.

2.1.1 A brief historical overview of aggregation of judgements

The first historical account of 'aggregation of judgements or opinions' was found in Sir Francis Galton's work on 'Inquiries into Human Faculty and its Development' in 1883. In this notable work, Sir Francis provided a statistical reasoning to the aggregation of judgements' framework. The experiments conducted by him paved the way for the statistical justification of an increased sample size; where a large crowd was asked to find a solution to a problem which may or may not have a true answer. The collective judgements of the crowd would often get close to the true answer, and in some situations even predict the true value with accuracy. It raised some important questions. One of the main issues surrounding the study of uncertainty around aggregation and its impact lay in those situations where individual groups neither know the answer nor have any idea about what could possibly be the answer. This then fed into a more interesting question that when the individual judgements are known, what additional information can the aggregated result provide? One of the most challenging questions was whose judgements would one end up with once such judgements are aggregated? The collective judgement is a reflection of an individual in the group, but, since this is aggregated by the decision maker or the statistician, then, the question arises whether his or her judgements would make an impact on the total group's judgement. Furthermore, would the statistician's view of the group and also the prior knowledge influence the aggregate view? The answers to most of these questions do not exist.

In the work of Sir Francis like in most other case studies surrounding the aggregation literature, the true value of the underlying parameter of interest is often known to the decision maker (Winkler and Clemen, 1991). But, this is not uncommon that is not a known or agreed method to reveal the unknown for which it sometimes become impossible to search the true values of the interest quantity. Therefore, the scientific community could never really provide a justification as to why or how the crowds and groups got the right answers to problems that have no known or true answers. A way of understanding the phenomena of this wisdom to find out the true value is to analyse the aggregation of the opinions of the people in independent and large groups so that it may lead to error cancellation. Therefore, all that remains will be the information provided by each individual in the group, and when that information is aggregated by simple averaging, it often turns out to be close to the correct answer or a perceived correct answer (Galton, 1883). The famous example coined by Galton is a proof of this. As mentioned in chapter 1, in an experiment he had asked a group of lay persons to judge the weight of an ox and then on averaging the weights, he concluded that the correct weight was exactly same as that predicted by the crowd. However, disastrous decisions with aggregated judgements (Surowiecki, 2005) have not been ignored in the literature. The reason for such mishaps revolved around the fact that no one in the group had any idea about the true value of the unknown. Not many research articles are made available in the public domain that highlight the failure in decision making based on aggregation of judgements (Surowiecki, 2005). On the contrary, in situations when a group that is responsible for making decisions comprises people who have ideas around the true answers, the group decision should invariably lead to the correct decision or approximately close to the right decision (Surowiecki, 2005).

Following Galton's work (1883) and through the later works by Kahneman and Tversky (1982), the use of Bayes' rule for aggregation has been a prominent rule or heuristic that has been used by groups for aggregation of judgements across several situations and different disciplines. Kahneman and Tversky (1982) stated that the Bayes' rule offered people a set of rules that helped them to make decisions and also helped in incorporating the new information as and when they become available.

2.1.2 Background to experts, judgements and psychology behind judgements

Introduction

Well defined data often draws on data sets collected through survey questionnaires or data collected through observed experiments or historical records (Vose, 2008). These kinds of data are not always readily available and are also not suited to many inferential analyses (Vose, 2008), the prime reason being the subjectivity and uncertainty that surrounds the real life problems. If one were to take a step back and reflect, then the amount of experience a person has in performing a task would hold a far more in-depth understanding of the job, as opposed to generating data under controlled experimental conditions towards understanding that job. However, quantifying the experience in a meaningful manner is often challenging. Although having successfully achieved the task of quantification, the judgements could then act as a source of data, and provide a better insight into the understanding of the uncertainty surrounding the problem at hand. This would, in turn, enhance the process of decision making. The need to look beyond well-defined data often pushes one to explore the idea of judgements or opinions of experts, and this kind of collective judgements often provided an alternate source of data. This then very rightly raises two key questions; one as to who would one term as an expert; and second, how many experts would one require for decision making? The second question on the number of experts would rightly be governed by the problem at hand. However, these two questions naturally draw various other issues that surround experts' judgement. Discussion of these issues in an intricate detail is beyond the scope of the thesis; however, the psychological aspects surrounding expert judgements, which cannot be entirely ignored, has been briefly discussed subsequently in this chapter.

The expert

There are various definitions of 'an expert' in literature. An expert is defined as 'a person with substantive knowledge about the events whose uncertainty is to be assessed' (Ferrel, 1985). Experts may also be defined as people who have substantial knowledge of the research area that is being studied (O'Hagan et. al, 2006). Otway et. al (1992), referred to an expert's state of knowledge at the time of response to the question. Further, O'Hagan and colleagues, (2006) stated that not only is it necessary for an individual to have the knowledge but the individual's ability to also use that knowledge is what determines his/her expertise and classifies him/her as an 'expert'. The question on how many experts need to be consulted incorporated an interesting statistical debate. Broomell and Budescu (2009) have stated that the need and desire to capture the maximum amount of information often leads the decision makers to seek advice from multiple experts, and aggregate their opinions so as to achieve more accurate decisions. Shirazi (2009) stated that experts' judgement are considered uncertain, hence having multiple experts forms a more inclusive data source of information around the problem. The use of multiple experts can also be viewed as a source of increasing the sample size and obtaining as much information as possible (Winkler, 1921; Vose, 1996). Quigley and Walls (2010) have summarised the issues surrounding experts' judgement into two broad categories of eliciting and using the judgements, further includes a selection of experts, determination of expert panel size, ascertainment of calibration and aggregation methods.

Subjective aspects underlying judgements

Hogarth (1987) stated that the human or subjective component behind experts' judgement cannot be entirely ignored, although these are almost impossible to assess and/or measure in the context of pure statistical and mathematical research. Psychology literature and research further suggest that humans tend to have directional biases and that if given a task of identifying numerical and alphabetical serial orders, memory processes work differently, but in both the cases the middle term or the multi-term of an array is identified most slowly (Jou, 2003). In most of the expert judgement elicitations, the experts are often asked to assess their middle quartiles while specifying probabilistic judgements (Cooke, 1991; Jouini and Clemen, 1996; Quigley and Walls, 2010), which offer an interesting insight into how well aware the experts are in specifying these quartiles. Although every expert elicitation task follows standard protocols to reduce bias but due to the very nature of human psychology, it does lead one to wonder whether such judgements are usually directionally biased? Psychology literature further reveals that the mental representations and type of questioning are likely to sway inferences in probabilistic and statistical domains, and the reasoners do not always correctly estimate conditional probabilities (Gong et. al, 2010).

Moreover, it has been seen that the people can solve problems where they can provide reasons extensionally, in spite of whether or not they have been able to process the information like the way a human mid is adapting to solve issues. Reasoning about initial probability is likely to be affected by the structure of the problem and the form of the question asked. It appears that human cognitive architecture runs some computationally trivial algorithms, like Bayes' rule, for drawing probabilistic and statistical inferences. As mentioned by Cooke (1991), there might arise some differences in the way the middle term is characterised or a way in which the conditional probabilities are perceived as most of the experts' judgements are specified as quartiles or percentiles or as conditional probabilities. Jou (2003) specified that this makes mathematics an interesting subject to assess the behaviour of the upper and lower quartile values in relation to the median or middle values specified by the experts. It is believed that both the developmental and individual differences in reasoning can be at least be partially explained by differential access to knowledge stored in the long-term memory of a human brain (Ben-Arieh et. al, 2006).

There are several steps undertaken during the expert elicitation process to make sure that the bias is reduced from the model. At the same time, this might minimise the bias which occurs when the experts have been provided with some estimates from another context. This bases his or her opinion on the estimation to assess other variables. This bias may be termed as an anchorage, and a more formal definition of an anchorage may be viewed as an occurrence where the experts elicit probabilities of several events with respect to an initial assessment (Quigley and Walls, 2010), thus rendering a pattern in the assessments. Checking the judgements for the source of potential patterns can always be a pivotal task due to the directional bias that exists within the human psyche. There might be other complex cognitive processes and executive function at work which is beyond the scope of this study, which underlies the expert elicitation process. It nevertheless makes one wonder whether such cognitive concepts possibly underlie even more complex mathematical reasoning and numerical problem-solving. Also, as individuals tend to think differently, and some ways of thinking resist change even after the substantial amount of training and instruction; this can perhaps influence the way one learns and perceives mathematical and scientific concepts (McNeil et. al, 2010). However, experts with shared knowledge coming from similar educational backgrounds and exposure to similar work environments may still differ in the reasoning and judgements (Winkler, 1999). Disagreements or differences in opinions and judgements of the same variables are bound to occur, especially because of the differences in individual understanding, the perception of views, and memory (Ben-Arieh et. al, 2006). Another possible source of difference can be demonstrated by memory and cognition tests, where the results show that complex and simple tasks involve two distinct learning systems (Lee, 1995), thus making the difference in opinions on the same quantity, evident.

2.1.3 Dependence and probability axioms in aggregation

Dependence or correlation plays an important part in the aggregation of judgements. Two critical ways of modelling dependence i.e. Bayesian and axiomatic, have been found in the literature. In the Bayesian setting, as commonly used and described later in this chapter, the estimation of the dependence parameter is a challenging task, barring the conditions where the distribution reduces to bivariate normal (Chhibber et. al, 1992). In the work by Jouini and Clemen, (1996), the dependence parameter is modelled using copulas, which gives more flexibility to the dependence modelling between the experts. In the use of copula models, the simple combining rules fall out as special cases of these models (Winkler, 1999). The decision maker's posterior distribution is then defined as $P(\theta|f_1...f_n) \propto L_c \times [(1 - F_1(\theta))...(1 - F_n(\theta))] \times \prod f_i(\theta)$ where L_c represents the copula density function and $P(\theta|f_1...f_n)$ gives the posterior distribution. The likelihood function is expressed in terms of the copula and the marginal or prior distributions.

A copula function provides a way to write the joint distribution function given the marginal (Nelsen, 2006). In this approach, the individual judgements are entirely different from the judgements about dependence. The approaches for calibration are data based and involve only the marginal distributions. The dependence between the experts is encoded in the copula function. Hammitt and Shlyakhter (1999) implemented the Clemen's copula model so that it can aggregate the experts' opinions to study the global climatic changes and cancer risks. But, the experts' judgements have been studied as the marginal while combining the dependency between the judgements using the copula function. This has helped in further development of the model to produce a posterior distribution through the formula discussed above. Kallen and Cooke (2002) proposed a copula-based aggregation model for expert judgements as a follow-up study to Jouini and Clemen (1996), where they explored the limitations of the latter study and suggested further exploration of the mathematical construct of the copula.

Apart from the Bayesian aggregation methods, the copula method also allows for modelling experts dependence. The copula method interprets the expert judgements as data and uses them to update the decision makers' prior distribution for the target variable to produce the posterior distribution. The copula approach provides a way of isolating the description of the dependence structure from the individual factors and provides a technique that allows the combination of marginal models with a variety of possible dependence models, and to investigate the sensitivity to the dependence specification, which is also suggested by Kallen and Cooke (2002). In axiomatic approaches to aggregation, difficulties often arise. These are centred on the probability axioms for aggregation of expert judgements, such as preservation of independence, monotonicity, and continuity. These properties were earlier proposed by Morris (1974), and have been further studied by French (1986) and Genest et al., (1986). Ironically, all these properties have shown to fail in the paper by Lindley (1986). The main reason for the failure of these properties has been attributed to the limited evidence available, which also goes to show that aggregation using simple averages of the multiple judgements as a model performs equally at par with other mathematical aggregation models (French, 1986). Mixture distributions and copula models are both based on modelling statistical dependence between experts judgement, and use multiple experts to assess the same variable. Though the copula approach fulfils the criteria of aggregation in terms of exchangeability, it overlooks the axiom of external Bayesianity (Wisse et. al, 2008). In addition to this, Jouini and Clemen (1996) chose the dependence parameter value based on the decision maker's belief. Due to the modelling assumptions of zero cross correlation, and positive dependence, the copula modelling approach to aggregate expert judgements is not a suitable modelling framework for this thesis. However, modelling dependence using copulas could possibly be useful if negative correlations and cross-correlations could be incorporated into the model. This is one area which needs further investigation and the reasons can be attributed to the unanimous choice of modelling dependence using copulas (Nelsen, 2006; Jouini and Clemen, 1996; Kallen and Cooke, 2002).

2.1.4 Aggregation of expert judgements

Aggregating or combining the data leads to a reduction in errors (Armstrong and School, 1985; Hogarth, 1987), especially because humans tend to ignore the dependence between variables, and also because the mind introduces its own complications and perspectives. Though formal elicitation allows the collection of different points of views and perspectives from experts with different education and cultural background (Zio and Apostolakis, 1997), it would even be adequate for decision makers to obtain the individual probability assessments. It further helps in making decisions in context to individual assessments other than attempting to integrate these to produce a single distribution. One could then argue that this would rule out the purpose of aggregation of expert's judgement. However, the reason for aggregating judgements remains an on-going debate because a clear and concise view of the experts' is captured with individual judgements, yet the aggregated result may or may not have such an interpretation (O'Hagan et. al, 2006). Evidence in psychology literature of decision making (Hogarth, 1987; Jou, 2003), suggests that humans have difficulty in combining information especially from dependent data sources and they have difficulty in specifying a median value or a middle value (Jou, 2003) in which case aggregating the judgements becomes crucial as the experts may provide median value as judgements or can have different perceptions to the questions that are being asked.

Aggregating the probabilistic expert judgements would be useful especially when there are multiple experts on the same area of expertise, and multiple experts judging the same scenario, aggregating the judgements into a single distribution so as to capture the similarities in their perspectives on the situation and/or differences in their opinions, that can further add to the differences in their perspectives, understanding and thoughts on the questions related to the same situation (Winkler, 1981). Further, Arrow's impossibility theorem of social choice (Fishburn, 1970) states that a consensual solution derived from all experts judgement collected is a socially acceptable solution when they are all assessing the same thing or in other words, the individuals have some commonality between them (Yan et. al, 2011). Although the social choice that drives the consensual decision deals with utilities that are assessed by experts which are different from probabilistic judgements; the underlying key aspect stresses on the need for aggregation so as to arrive at a meaningful decision. When experts are pooled together, collectively they offer sufficient insights leading to the building of a comprehensive theory and support the process of decision-making.

The judgements that are given by the experts' often rely on the use of known strategies, and the experts tend to rely more on the knowledge that they follow through a routine and a process that leads to a particular knowledge, thus making the judgements conditioned with respect to all relevant information. For example, if an expert starts with the event most likely to occur and assigns it a probability value and makes all other probability values relative to this one, then these judgements become conditioned and introduce biases in the model (Goldstein and Hogarth, 1997). Also, Bonabeau (2003) stated that one should not trust one's intuition about a process or a system, rather one should adopt a radical way of thinking. This then draws the argument on experts' judgements that can then these judgements be postulated as to how radical the thought process is behind the routine knowledge? This draws an attention on the arguments made by the experts' judgements to understand whether the judgements can be postulated and proved as radical in context to the thought process behind the routine knowledge.

The approach of thought can possibly be attributed to one's experience, education, and consistently sound and rational choices through life. Basic differences between the factors affecting the thought process possibly result in disagreements or differences in opinions between experts on a same scenario or situation (Goldstein and Hogarth, 1997), which is of interest to the decision makers for they can then account and justify this difference with a combination of different opinions and judgements (Moon and Kang 1999). Also, mathematical aggregation makes the analyses auditable (Hanea et. al, 2017).

2.1.5 Behavioural aggregation methods

There are potential benefits of different modelling approaches to combining experts' judgement, which is broadly categorised into behavioural and mathematical models. The issue over mathematical aggregation versus behavioural aggregation is also an on-going debate and there is no evidence in the present literature to support that one is better than the other (O'Hagan et. al, 2006). In mathematical aggregation, the experts do not influence each other's decisions or subjective probabilities. In the behavioural approach, the experts are allowed to share their judgement and reassess their distributions. Some of these techniques are known as the Delphi method and the Nominal Group Technique (Ferrell, 1985), which are discussed later in the chapter together with the relevant mathematical aggregation methods to this research in detail. If expert judgements are used to assess uncertainty within the behavioural models, it does not necessarily imply that just because individuals may perform poorly while assessing uncertainty, group elicitations should not be done. At the same time, it also does not mean that complex mathematical models should not be explored which may offer a better tool to the understanding of the aggregation while assessing the uncertainty of the variables (O'Hagan et. al, 2006).

The aggregated judgement would help the decision maker in making an informed decision encapsulating the individual behaviour and the dependency between experts' if any (O'Hagan et. al, 2006). The statistical model of aggregation of experts' judgements has a better power of the test as opposed to individual judgements. By increasing the number of experts there is an increase in the statistical power of the model based on linear pooling of judgement means, which can easily be explained due to an increase in the sample size; and the aggregation itself helps in making an informed decision around the optimal number and choice of experts (Hogarth, 1987). Another view was presented in a survey paper on judgement aggregation by List and Puppe (2009), where the importance of aggregating experts' judgements was reviewed. It was also stated that the collective or aggregated opinion is merely an outcome of a collective thought and action process (Bonilla, 2006; List and Puppe, 2009). In the behavioural aggregation methods, approaches are required to somehow combine the experts by making them interact in some way or the other. These are done by either face-to-face group meetings or sharing information via virtual methods and other specific procedures such as the Delphi method or the Nominal Group technique (O'Hagan et. al, 2006). In the Delphi method, the experts first assess and present their probability distribution and then a group discussion takes place, after which the experts may or may not update their probabilities. The Delphi method requires mathematical aggregation of the final probability distributions (Okoli and Pawlowski, 2004).

Studies of different researchers like Gigone and Hastie (1997) have compared the mathematical and behavioural aggregation methods but these have been consistent with the mathematical models which are simple statistical methods of averages and linear combination. Therefore, the results from these studies have been mixed, with some advocating the behavioural aggregation methods and some the mathematical ones (Winkler, 1991; Shirazi, 2009). Several studies on the behavioural aggregation of expert's judgements have proposed the use of fuzzy theory and linguistic criterion's at multi-granularity levels (Ben-Arieh et. al, 2006; Salo, 1995; Vanicek et. al, 2009). Cornell (1996), proposed aggregation of expert judgements at two levels. At first, he combined the experts' judgements by simply averaging and providing each expert an equal weight. Secondly, he sent the results back to the experts so that they get an opportunity to revise the judgements depending on what others would say. This technique is based on the Delphi method of aggregation of judgements, which is a behavioural method of aggregation.

Gigone and Hastie (1997) showed that simple mathematical average of individual experts' judgement outperformed the group judgements in their experimental study of groups for economic and policy decision making within the context of democratic institutions and societies. Flores and White (1989) showed in their experiment on comparing mathematical and behavioural aggregation methods that both the methods i.e. behavioural and mathematical were at par in their performance for the aggregation experiment with respect to group judgements. The challenge, therefore, was in finding mathematical modelling procedures which also involved behavioural aspects for aggregating experts' judgement. Overall, the scientific community stands divided on the aggregation of experts' judgements, and there is an on-going debate surrounding the different approaches to establish a 'best' known method (Winkler, 1999).

2.1.6 Mathematical aggregation of expert judgements

The two commonly used methods in the mathematical aggregation of expert judgements are the Bayesian methods and the axiomatic approaches (Winkler, 1999; O'Hagan et. al, 2006). Many different methods are adopted for the axiomatic aggregation; some of them are detailed in the paragraphs below.

Cooke's Classical Method for aggregation and linear pooling

Cookes' classical method (1991) is one of the widely used mathematical aggregation models for experts' judgements. This method works on the linear pool of weights of experts' opinions using seed variables or real data. The experts are given a set of questions to give their opinion on, the answers to which are already known to the decision maker. The experts are required to specify their probability in terms of quartiles and a score is assigned to weigh the judgements. The scoring rule differs from information and calibration. Based on their performance on the pre-defined exercise, weights are assigned to each expert, and the aggregation is a simple weighted average of the opinions. These weights are a combination of the information and calibration score.

The potential criticisms of Cooke's classical model are that it ignores the experts correlated errors in judgement, and it may happen that two experts potentially perform equally well and are assigned the same weight but in essence, one expert can be made redundant, as he/she may be highly correlated with the other. So in essence as the correlations are ignored; a final outcome is a large number of experts who are highly correlated, but the information received in form of their judgements are similar; hence having a large number of experts does not necessarily help as there is no additional contribution to the knowledge base of the decision problem.

Lin (2011) studied the aggregation by Kullback-Leibler divergence criteria of uncertainty judgements in risk assessments to evaluate the expert's knowledge on particular questions. This has helped to develop a new model which would better rank and assign scores to the experts, those who are in line with the Cooke's model. Along with this classical method, linear pooling of opinions was also used to aggregate opinions in a Bayesian framework. Some of these initial models were proposed by Genest et al., (1986).

Bayesian method of aggregation

The first Bayesian method of aggregation of expert judgements was offered by Morris (1974, 1977). Since Morris's novel work, many Bayesian methods have been developed and introduced over the years. The Bayesian method is one of the methods that allows to model dependence between expert judgements (Chhibber et. al, 1992). The problem of dependence is one of the major issues in aggregation, and being difficult to evaluate, it is also the central part of a combination of expert's judgements (Kallen and Cooke, 2002). Judgements from multiple experts about a parameter are extremely informative if the experts are statistically independent but, when there is a statistical dependence between the experts then the judgements are correlated, and the relevance and usefulness of information decrease Broomell et al. (2009). Therefore dependence modelling forms a key issue in the aggregation of expert judgements.

In the Bayesian aggregation framework, each expert's distribution is taken as a prior probability distribution, which is then multiplied by the likelihood of the occurrence of the event in the light of the data to give the posterior distribution. The likelihood function of the Bayesian aggregation model allows the flexibility to calibrate the experts' judgements and also accounts for inter-expert dependence. The systematic location biases for the experts' can be modelled either as an additive model or a multiplicative model.

In the paper by Chhibber et. al, (1992), assuming, $P(X|E) = P(E|X) \times P(X)/K$, where the $P(X|E)$ represents the decision maker's belief of X in the light of event E, where, E is the expert judgement; $P(X|E)$ is completely determined by $P(E|X)$, which is also known as the likelihood function, K being the normalising constant. For example, if there are two experts and under the assumption that the location bias model to be multiplicative, the likelihood function can be viewed as a bi-variate log-Normal distribution having the following form:

$$L\left(x_1, x_2|x\right) = \left(\frac{1}{2\pi x_1 x_2}\right)\exp\left(\begin{array}{l} \frac{-1}{2(1-\rho^2)}\frac{(\ln(x_1)-\ln(x_2)-\ln(b_1))^2}{{\sigma^2}_1} \\ +\frac{(\ln(x_2)-\ln(x)-\ln(b_1))^2}{{\sigma^2}_2} \\ -\frac{2\rho\left(\begin{array}{l}\ln\left(x_1\right)-\ln\left(x\right) \\ -\ln\left(b_1\right)\end{array}\right)(\ln(x_2)-\ln(x)-\ln(b_1))}{\sigma_1\sigma_2} \end{array}\right)$$

where b is the median of E and ρ is the correlation coefficient; the inter-dependence between the experts is quantified using ρ. Thus the median of the posterior distributions can be computed. Under the circumstances mentioned above and with the prior being modelled as log-normal the Bayesian aggregation then reduces to weighted geometric mean (Chhibber et. al, 1992). One of the critical issues surrounding Bayesian aggregation methods is the specification and choice of the likelihood function (Winkler, 1999). In the example above, the priors or the marginals are assumed to be Gaussian, and hence the likelihood form is bivariate Gaussian, but with an increase in the number of parameters for assessment, the Bayesian aggregation method becomes non-trivial and mathematically cumbersome (Chhibber et. al, 1992).

Another method of aggregation within the Bayesian framework of aggregation is the Mendel and Sheridan's model (1989). This model uses non-normal probability distributions, assuming that each expert provides m fractiles of his distribution and that each expert's actual outcome is assumed to fall within (m + 1)n bins. This approach provides joint calibration, as it produces probability distributions that are based on a multivariate setting but this method also requires the estimation of the parameters of the likelihood function, which is challenging in practice (Shirazi, 2009). Although the decision maker can assess the prior distribution from the data subjectively, this may be a difficult task, especially if one considers the Mendel and Sheridan's model, as one is then required to assess elements of the covariance matrix of the probabilities associated with each cell in the (m + 1)n array.

Overall, identifying an exact likelihood function for expert probability is considered challenging and difficult within the Bayesian aggregation of expert judgement (Chhibber et. al, 1992; Morris, 1977). However, there are studies that have computed likelihood functions where the distributions are normal (Chhibber et. al, 1993; Morris, 1977; Winkler, 1999).

Moments Method for aggregation

Wisse et. al, (2008) studied a method of moments for aggregating expert judgements based on Bayes linear methodology. The objective of the paper was to be able to aggregate expert judgements in a non-Bayesian way using expert assessments of moments. The authors used the extended Pearson-Tukey method to derive assessments of the first and second moment for the quantile assessments. This method ignores the computational complexities of continuous probability distributions.

In addition, Genest et. al (1986) developed a method of aggregation of judgements where the experts just specify certain moments of a distribution rather than specifying the whole probability distribution. Zio (1996) and Xu (2000) studied the aggregation problem with analytical hierarchy process. Xu (2000) looked at the weighted geometric mean of the analytical hierarchy process for aggregating expert judgements. This method works with an assumption that the experts' judgement matrix is consistent and the method works perfectly well within the set criteria. Zio (1996) used the simple weighted averaging method within the hierarchical structure. The hierarchy is built based on the confidence that the decision maker has the experts and these weights are further used to aggregate the judgements.

Maximizing Entropy method for aggregation

Myung et. al (1996) and Wood et. al (2006) studied the aggregation of expert judgements by maximising the underlying entropy function. Maximum entropy method is a statistical inference procedure that satisfies the five axioms of invariance, uniqueness, null information, system independence and subset independence. The inference gives the best probability estimate from the given information without assuming any other knowledge beyond the constrained set of information. This method of maximum entropy stands sound not only on theoretical grounds but also retains the versatility to be used in different practical applications. In this method, the event to be predicted and the individual predictions are assumed to be discrete random variables. The Shannon's information is used to derive the aggregation rules for combining two or more expert predictions into a single aggregated prediction that approximately calibrates different degrees of expert competence. However, the problem of modelling the dependence between the judgements remains unanswered with this method.

Other models in literature

Zio and Apostolakis (1997) proposed a mixture distribution to aggregate the expert judgements by studying within expert and between expert variability. This is one research which has addressed the issue of aggregation of between expert and within expert judgements. However, the work has not been extensively used in practice. The authors stated that the main motivation behind developing this paper was mainly due to the aggregation models that tend to produce composite distributions. This is to reflect the less uncertainty as opposed to an expert's individual distributions. And one of the possible reasons for this discrepancy is the attribution between the expert's variability that is not modelled and captured accurately. The difference or similarity of experts' assessments on the variables is modelled but the dependence between the experts is often not captured accurately. Some of the reasons for this dependence may be attributed to shared information, shared knowledge and similar education background among other factors, as already discussed earlier in this chapter, but these between expert dependence is often not modelled. This study attempted to account for between and within expert variability in the context of the future climate of Yucca mountain vicinity. However, the theoretical framework could not be applied due to the constraints of equal weights of experts and the assumption of the underlying probability distribution of the unknown parameter.

Jouini and Clemen (1996), considered a Bayesian decision maker who is interested in making an inference about an unknown parameter say θ. The decision maker defines his prior probability density as $p(\theta)$, and n experts provide their opinions $g_i(\theta)$. Each of these g_i are the expert's personal probability distribution for the parameter θ. Then by using the Bayes' theorem, the posterior distribution is given by $p(\theta|g_1,...,g_n) \propto p(\theta) \times f_n((g_1,...,g_n)|\theta)$, where $f_n((g_1,...,g_n)|\theta)$is the likelihood function for the expert judgements. Following from Clemen and Winkler (1991), Jouini and Clemen, assumed that for the aggregation problem, everything that the decision maker knows is incorporated into the experts judgements. For this reason, the prior adopted in their approach was an non informative prior density.

If we take H_i and h_i as expert $i's$ cumulative distribution function and density function respectively, then the likelihood function is given by

$f_n = L_{c_n}((1 - H_1(\theta))..., (1 - H_n(\theta)))h_1(\theta)...h_n(\theta)$

where L_{c_n} is the copula. The dependence value is captured through the correlation value ρ and Gaussian copula parameters are estimated for the model.

The reasons for the choice of Gaussian copula can be attributed to the properties that are mentioned by the authors in their paper; such as symmetry and exchangeability while accounting for the statistical dependence between expert judgements. Jouini and Clemen (1996) proposed a Frank copula, for the estimation of posterior probabilities and benchmarked the results of Frank copula to the Gaussian copula.

Frederic et. al, (2012) proposed a product of odds model to aggregate expert judgements. In their model, the statistician or the decision maker is able to incorporate his beliefs into the aggregated assessment of the expert judgements. The underlying assumption of their model is that the decision maker has an access to the information on the unknown event or more precisely the decision maker has his own beliefs around the unknown parameter that can be incorporated. This model also ignores the dependencies that might exist while consulting more than one expert. This model stresses on the combination of the statistician's knowledge and the expert's knowledge, which in essence ignores the situations where the experts may or may not be known to the statistician.

2.2 Summary

This section provided an in-depth literature review of aggregation of expert's judgement. It provided background into aggregation and the different types of aggregation procedures and methodologies that exist in literature; incorporating both behavioural and mathematical models. The relevance of aggregation of wisdom or knowledge of the crowd was justified and the issue that was highlighted lay in successfully identifying the definition of a 'crowd'; leading to a definition of an expert. This further provided a background to the psychological aspects that surround the process of decision-making across disciplines. The psychological aspects shed light on the fact that humans use rules and heuristics to provide a solution to problems throughout history. The Bayes theorem has always had a prominence within the aggregation framework across disciplines (Surowiecki, 2005). Psychologically, knowingly or unknowingly, people have always used the Bayes theorem to come up with solutions to unknown problems using their beliefs. The Bayes theorem is viewed as an intuitive mechanism where people update their judgements or beliefs as new facts become available. As a result, they can reach a more meaningful conclusion (Kahneman and Tversky, 1982). The existence of correlations between experts especially when they have shared beliefs and work environment, among other factors were highlighted.

Therefore, in an attempt to gather and analyse a coherent dataset, which comprises these judgements as probabilistic data the dependencies and correlations would undoubtedly, play a critical role. On the other hand, the mathematical aggregation methods consisted of processes and analytical models which helped in using individual probability distributions to produce a single 'aggregated' probability distribution (Winkler, 1999). At the same time, the behavioural aggregation can be viewed as the combining experts being opposed to amalgamate the probability distributions. Behavioural method attempted to generate agreement among the experts by getting them to interact in some way, based on the assumption that as the information got shared through these interactions, better arguments and information was likely to influence the group and that redundant information was likely to be discarded, thus having the final decision practically in agreement with all the experts (O'Hagan et. al, 2006).

However, judgements from experts would be invariably correlated due to their shared experiences and other psychological factors. Therefore, while aggregating the judgements mathematically, there is a strong need to model the dependencies that exist. Aggregation of the individual subjective distributions to form coherent unifying distributions would enhance the understanding of the unknown parameter, whose true value can never be measured or known. There might arise circumstances, where the decision maker is dealing with a secondary dataset, incorporating his/her understanding on either judging the credibility of the experts or on the unknown quantities are next to impossible. In such situations, making use of the available data would be the most sensible way forward. As stated by Winkler (1991), in the light of uncertainty, all available information must be aggregated to gain a better understanding of the problem that is being dealt with. Further, in all the mathematical models that have been discussed in this chapter, there is no unifying mathematical model for aggregation; more rightly so, as each model is dependent on the problem that it is being used to solve. In a comparison study on the performances of the mathematical models of aggregation of expert judgements, by Hammitt and Zhang (2013), it has been observed that the equal-weight combination rule, which is the method most often applied in practice, has the worst performance. The simulation of the data for this comparison study further stated that the assumptions of positive correlations between the expert's errors are not a likely condition to be found in practice. While discussing the dependence modelling method with copulas, it was stated that the decision maker concentrates probability on values to which all experts assign significant probability and very little probability to values to which any expert assigns a small probability or a zero probability.

In summary, the literature provided a gap that exists with regards to mathematical models for aggregation where the judgements are correlated. Having stated this, there exist models that do study dependencies between judgements but these have a narrow focus, primarily concentrating on either positive correlations or focussing on a scenario of consulting a maximum of two experts. It also provided evidence for the lack of methodological approaches encompassing both the research philosophical paradigms which can also incorporate a wide range of data structures together with addressing the issues of correlations.

2.3 Review of the proposed mathematical methods

2.3.1 Introduction

This section provides a literature review of the proposed methods i.e. non-parametric and empirical Bayes methods. These methods are further developed for aggregation of correlated experts judgements in this thesis.

2.3.2 Non-parametric method

Background

The development of non-parametric methods was relatively slow until the end of the Second World War, but since then their growth has touched almost every phase of statistical activity (Hjort et. al, 2010). There has been a prominent statistical inferential procedure throughout the history, but, the aspect of inference surpasses the parametric and non-parametric forms of the statistical inference. At last, this helps in incorporating both as noted by Kass, (2010) and Tukey (1960). Furthermore, the need for data analysis beyond the known parametric forms has been a prominent feature (Tukey, 1962). Tukey advocated the need for non-parametric methods as the first steps towards providing more realistic frameworks for understanding and analyses of data. It is also well established that the study of random patterns of data is often guided by a set of rules and a set of assumptions. These assumptions are often challenged and argued across disciplines, and have helped in the formation of the norm that most mathematical models do help in the understanding of the complex phenomenon of the real world, not necessarily providing an in-depth solution (Vose,2008; Alejandro et. al,2011). Another reason for the advocacy of non-parametric methods by Tukey was the presence of fluctuations in the data with a reasonable distribution but unlikely to fit a normal distribution. The need to use data for both exploratory and confirmatory analyses has been stressed upon, and it was believed that modifying the data to fit a statistical model is far less credible than using the data to tell its own story (Tukey, 1960). The use of five measures; namely, the median, quartiles and the two extremes; the maximum and the minimum, were popularised for understanding data and helped to reiterate the clear advantage of non-parametric forms over parametric distributions. It was obvious that these measures would be always defined for all empirical distributions and with the use of exploratory data analysis, Tukey (1962) also tried to provide means to reduce the errors in formulating the statistical models for further drawing inference.

There are several other aspects that do contribute to the lack of structure and understanding of these replica models of reality. As a result, the assumptions surrounding the parametric model fitting to understand the data are often relaxed. The statistical decision theory and inference procedures are built on the parametric forms of probability distributions and provide a wide range of tools and techniques to deal with data structures; however, within the realm of these decision theoretic frameworks, the question that is often challenging is to have a procedure or a set of techniques that can capitalise on the nature and structure of data that are not well defined. A host of data structures would then get a representation and meaning through the procedures that are not primarily dominated by the probability distributions (Tukey, 1960; 1962).

One of the major drawbacks in using non-parametric statistical methods is the asymptotic inferential procedures that underlie these statistics. Although, non-parametric methods have fewer assumptions on the data, the methods lack power for a small sample size. Therefore, according to the literature, unless the normality conditions of the underlying data are completely violated, it is not advisable to use non-parametric methods. These methods are known to obey consistency where the posterior distribution properly accumulates its mass around the true model with an increased sample size. However, for an identical problem set, this accumulation has a different rate of convergence in comparison to the parametric methods (Hjort et. al, 2010).

Aggregation of judgements using non-parametric method

In the context of aggregation of experts' judgement data, several non-parametric inferential procedures exist in literature. However, most of these methods have been used for ranking the expert judgement, and providing orders to the judgements in social decision-making context (Muller and Mitra, 2013; Kauko et. al, 2003; Tsyganok et. al, 2012). One of the popular non-parametric aggregation approaches is that of maximisation of entropy (Myung et. al, 1996; Wood et. al, 2006). This method makes use of the five point estimates as described in the previous paragraph of this chapter. Vincentization and parameter averaging are two other non-parametric methods used for combining judgements or combining information from independent sources (Genest, 1992; Gu, 2009). Both these approaches are fairly popular within psychology where the aggregation of information is of prime importance. It has been noted that the classical method of Cooke in respect to aggregate judgements has been compared with the non-parametric sign test, being developed by Clemen. In this context, the distance between the medians of the aggregated results and the seed variables were at first compared to measure the accuracy of the models. This study concluded that equal-weighting of experts was a flawed method at many levels and a more rigorous mathematical aggregated model was required to address the issue (Yin and Cheng, 2009). The allocation of weights to the judgements provided by the experts is a crucial task within the aggregation framework. Several mathematical constructs exists and some of these have been presented in this chapter. One of the most common construct towards weight allocation to the judgements is that of equal weights. Every expert's judgement is given an equal weight and it is then aggregated to provide a weighted average to the questions or variables under study. The allocation of weights becomes hard when the judgements are correlated. Cooke (1991), used a method where he tested the performance of the experts against questions that he had answers to, therefore, when the same set of experts were asked to assess and provide their judgements on an the quantity, whose values were not known, based on their previous performances, the weights were assigned to each expert. Also, in this approach, the author assumed the experts to be independent of each other. In a study of the Yucca Mountain range in Nevada, authors, Zio and Apostolakis (1997) and Jouini and Clemen (1996), in separate studies on the same subject, made use of the decision makers' belief, ranked the experts and assigned weights. Zio and Apostolakis (1997), used equal weightings in their mixture model method when they aggregated the expert judgements. Although Jouini and Clemen (1996), studied the correlation between the judgements, the correlation was assessed by the decision makers and not through the data. It was the decision makers' belief as to how he felt the experts were correlated and performed on the assessment tasks. Though the expert elicitation task is structured in a sense that it is aimed at reducing biases that might exists in the judgements, it is not entirely possible to control for the bias that might exist inherently within the expert. However, it is advisable to accumulate judgements from independent experts so as to gain maximum information but because inherent dependencies cannot be completely ignored in a multi-expert scenario, therefore discounting the correlated experts in terms of weight assignment of their judgements becomes necessary.

2.3.3 Empirical Bayes method

Background

Empirical Bayes (EB) method is a powerful data analysis tool (Casella, 1995) especially in the case of a multi-parameter estimation problem where known correlations or relationships between the different variables suggest pooling information across similar experiments gain a better inference of the underlying parameter. The Bayesian literature is predominantly concerned with the construction of the posterior distributions given the data and prior beliefs. Given the prior beliefs; which may be expressed as probability distributions, the construction of the posterior distributions would entail either maximise the information that is obtained through the assessments or minimising the errors that exist because of the discrepancies or misspecification of the actual and calculated parameter values. In EB, the method the prior and the likelihood are both estimated from the available data and then using the Bayes theorem, the posterior distribution is calculated (Efron et. al, 2002).

This method is applicable when the decision problem is presented repeatedly and independently with a fixed but unknown a prior distribution of the parameters (Robbins, 1956). Robbin's theory on EB justified the development of EB method to aggregating correlated experts judgement data. Bayesian techniques such as the EB approach, have also been used for addressing various practical issues over the years. EB methods are considered as powerful data analysis tools in present era. This is significant especially, when the inferential statements are made regarding the parameter which is based on data (Casella, 1985; Deely and Lindley, 1981; Carlin and Lewis, 1996). The problem that has been addressed in this thesis comprises drawing an inference from an observed statistical value upon the unknown value of a parameter and thus examining the chance of the inference to be correct (Mises, 1943).

Aggregation of judgements using EB methods

Empirical Bayes statistical inference theory could be divided into the parametric empirical Bayes inference and non-parametric empirical Bayes inference. The main difference between these two methods of inferences being that for the parametric empirical Bayes inference, the prior belongs to some class of parametric distributions with unknown parameters, which need to be estimated from the data. These parameters are also referred to as hyperparameters. The form of the prior in a nonparametric empirical Bayes inference procedure is not assumed to follow any known parametric form.

EB methods have been used in literature and across disciplines, and within the framework of decision analysis, these methods have been used for aggregation of various risk and other decision factors. Arbenz and Canestraro (2010) used internal data from the insurance institutions and other external sources of data to study insurance risk using empirical Bayes methodology. The prior distribution that was applied to the loss default models was assumed to follow beta distribution because of its suitability to the data structure. There are several pieces of evidence of use of empirical Bayes techniques for aggregation; for example, in the area of financial risk (Kiefer, 2006) this method has been used to combine expert judgements with observed data for the construction of low default portfolios of the various financial instruments. The prior distribution was modelled as beta distribution. It has been seen that in most of the financial risk studies, in particular within operational risk while using

EB methodology for aggregation of data, beta distribution is commonly used (Lambrigger et. al, 2007). The EB statistical technique has also been used for aggregating spatial data within epidemiological studies concerning disease mapping (Devine et. al, 1994) However, the epidemiological study does not account for any form correlation in the modelling.

2.3.4 Strengths and weaknesses of EB method

Like any mathematical method, the EB methodology has been criticised by several authors (Gelman, 2004) where the empiricism within Bayesian paradigm has been questioned. There is no denying fact that Bayesian inference has attractive features for its coherence and good frequentist properties (Petrone et. al., 2012). However, eliciting an honest prior might not be a simple task, and as a result adopting the empirical method where estimating the parameters from data would add more meaning and render a structure to the Bayesian analysis. In a fairly large sample, EB leads to similar inferential answers as a proper Bayesian inference. In a data-driven choice of prior hyperparameters situation, empirical Bayes is a preferred method of analysis as opposed to a fully Bayesian methodology (Petrone et. al. 2012). The inconsistencies within EB methods have been found where the issues lay with the decomposition of variance within the EB methods (Cooke, 1986). Within the context of this thesis, the decomposition of variance factor was important; and literature suggested that it was unreasonable that one expert always gave the same standard deviation for all questions and that EB methodology provided a first order approximation to reality. However, EB techniques were useful for estimating between expert variability for many sets of expert opinions. It was further argued that the question on the decomposition of variance might or might not be independent. This means the way applied by an expert in assessing the probability of one question might not be possible to be applied in another process or method. Hence, the dependence between the judgements across rows is not the same across the columns. Therefore, EB methodology may be applied to model the dependencies and aggregate the expert judgements.

A fairly common criticism that has always surrounded the EB methodology is the fact that the data is used twice to model; firstly, to decide on the prior parameters and then for model the dependence between the judgements. In this thesis, it is assumed that the judgements provided by the experts in terms of the point estimates come from a normal distribution and thus the prior is constructed from the data and is assumed to follow a normal distribution and then the likelihood is constructed using the same data. This prior and likelihood and then used to compute the posterior distribution. This is closely followed by further criticism of the property of exchangeability. In many real world contexts, exchangeability would not be a preferred property, and therefore the use of empiricism within the Bayesian context would be overruled. However, within the context of this research, the desired property of exchangeability does hold true. Exchangeability is defined in terms of the dependence structure similar to the definition in Jouini and Clemen (1996), meaning the joint distribution function is the same irrespective of any permutation of the judgements provided by the experts. In a situation like this, where the model is developed based on a real world data, the values of the parameter are unknown, hence deciding on a prior distribution such as a normal distribution for the data was a sensible thing to do. Moreover, the EB procedures provide evaluations over both the parameter space and the data space, and it is known as the plausible compromise between a strictly Bayesian modelling and a frequentist method towards modelling (Carlin and Louis 1999).

Another issue while dealing with expert judgements is the issue concerning the credibility of the experts. It can be argued that the experience and the rigorous training that the experts have been through justify their credibility as experts and therefore their judgements. However, in a more realistic setting, it might happen that the experts are biased but not credible. As a result of this, the estimates provided by the experts are different but these are biased estimates. But as the true value of the parameters remains unknown; the confidence on the aggregated result will not change irrespective of the results being small or large. The small or large results may be predominantly due to the positive and negative correlations that exist between the judgements of the experts.

2.4 Summary

This section summarizes the non-parametric and EB methods that are further developed to aggregate correlated judgements. The literature review provides an insight into the use of both these methods in aggregation but with a narrow definitive scope. It has been further established through the literature that dependence modelling plays a crucial role in aggregating judgements. Therefore, the gap in the literature on using these two methods for aggregation in presence of correlation sets the path for this research. Therefore, the thesis further aims to make a contribution in the field of existing literature regarding the experts' judgements aggregation while proposing the two alternative methods as mentioned earlier to aggregate the correlated judgements.

Chapter 3

Non-parametric method for expert judgement aggregation

3.1 Introduction

Since the experts are sought in absence of the true knowledge of the parameters or quantities of interest, the question around confidence on the expert's judgements may itself be challenging. The question essential to answer then is what weights are to be allocated to the experts on their judgements? This chapter provides a non-parametric model to aggregate correlated expert's judgements. This method, as stated in the literature review, makes no parametric assumptions on the judgement data and works with the first two moments from the data along with the covariance matrix of the judgements. Section 3.2 derives the non-parametric model for aggregation and illustrates the weight allocation to experts through hypothetical examples. Section 3.3 presents the identifiability of the parameters thus making inference possible. It is crucial property in statistics so that it is established that the parameters that are inferred can be estimated from the data. Section 3.4 discusses the findings of this chapter.

3.2 Non-parametric model for aggregation

The method developed in this chapter address the issue of weight allocation adopting a least square method minimising mean square error (MSE) using Lagrange multipliers. Usually, after the computation of the parameter estimates from the judgement data, the frequently encountered operation is to minimize the error that might occur during the estimation process. In order to minimize the errors, a quadratic form in terms of MSE is proposed. A quadratic form of loss function is preferred because the judgements provided by the experts could potentially be directionally biased. Although, it is assumed that the experts have provided the true values to the parameters and the aggregation is based on this assumption, it cannot be cross checked for the directional errors in the data unless the true values are known. Hence, the squared loss function is chosen over the other error functions. Several variants for solving a squared loss function appear in literature and when a function is to be minimized or maximized over a set of constraints, the standard way of dealing with it is by introducing a new set of constraints which holds the equation constant under the variations. This method is known as the Lagrange multiplier approach. When first introduced Lagrange multipliers were used to deal with problems with equality constraints because mathematics were predominantly synonymous with the study of equations (Rockafellar, 1993), hence the popularity of the multipliers were restricted to constrained optimisation with equality constraints. With the advancement in mathematics, the constraints were modified to represent the real world scenarios and hence the use of Lagrange multipliers slowly saw embracing inequality constraints. In order to find a local maxima or a local minima, the function was either minimised or maximised (was achieved by simply reversing the signs) depending on constraint or a set of constraints, dependent on the problem. A linear independence is assumed between the constraints that the function minimised or maximised is subjected to. It seemed evident that Lagrange only advised proceeding as if seeking a maximum or minimum, and that the key point being that the variables can be found by solving the given set of equations. Although, the validity of that assertion did not depend on the existence of an unconstrained extremum at the solution point. However, in today's context these multipliers are used with partial derivatives to locate a suitable maxima or minima (obtained by reversing signs) subject to a suitable list of constraints.

The Lagrange multipliers enable us to avoid making a choice of the independent variables and they permit the symmetry in a problem where the variables are entered symmetrically at the onset. These multipliers have been embedded in history for centuries. With a primary focus on being used for solving optimisation problems, named after the French mathematician Joseph Louis Lagrange, this provided an unique and accurate way of dealing with polynomials of a higher degree. These have had a rigorous use in the field of geometry where they have been used to study the impact between space and time at various different dimensions. Although the reach and usage of Lagrange multipliers stretch beyond the discipline of mathematics, restricting its use within the scope of this thesis, the usage is primarily dominated by finding the local maxima and local minima of the MSE of the aggregated judgements. The MSE in essence is computed as an estimator to estimate the error between the predicted and the actual or true value of the parameters (Taylor and Mann, 1983).

Using the mathematical construct of MSE and Lagrange multipliers, the theoretical basis for aggregation of correlation expert judgements is then expressed in terms of the following theorem.

Non-parametric method for expert aggregation

The following notation are used to define the theorem below :

Table 3.1: Notation used

z	represents the number of experts
n	represents the total number of questions
i, j	represent the experts where i and $j \in 1, 2, ..., z$
k	represents the question number; it acts as a counter for the number of questions, therefore $k \in 1, 2, ..., n$
X_{ki}	is the random variable representing expert i's assessment on question k
x_{ki}	is the realisation of X_{ki}
$\mu_{ki} = E\left(X_{ki}\right)$	is the expectation of expert i's assessment of question k

$\mu_i = E(X_{ki}\|\mu_{ki})$	is the true value for expert i's assessment $\forall\ k$
$\sigma_{ki} = \sqrt{E(X_{ki} - \mu_{ki})^2}$	is the standard deviation of expert i's assessment on question k
$c_{ijk} = Cov(X_{ki}, X_{kj})$	is the covariance between expert i and expert j's assessment over k^{th} question; where $Cov(X_{ki}, X_{kj})$ $= E((X_{ki} - \mu_{ki})(X_{kj} - \mu_{kj}))$.
C_k	is the covariance matrix for k^{th} question where c_{ij} is the $(i,j)^{th}$ element $\forall\ k$
C	represents the covariance matrix $\forall\ k$
C^{-1}	represents the inverse of the covariance matrix $\forall\ k$
Π	is the correlation matrix whose elements are σ_{ki}, c_{ij}, μ_k and ρ_{ij}
w_i	is the weight assigned to expert i
$\underline{w} = (w_1, ..., w_z)$	is the vector of weights assigned to experts
λ	is the Lagrange multiplier
$\hat{w_i}$	is the estimator of w_i
$\hat{\mu_{np}}$	denotes the non-parametric mean

Theorem 1 *Let the mean square error be defined as* $MSE = E\left(\left(\sum\limits_{i=1}^{z} w_i X_i - \mu_i\right)^2\right)$*, for* $k = 1$

Then the weights assigned to the experts' assessments that minimise the MSE *are proportional to the covariance matrix* (C) *as* $\underline{w} = \underset{w}{\operatorname{argmin}}\ MSE \propto C^{-1}$ *such that,* $\sum\limits_{i=1}^{z} w_i = 1$

Proof 1 $\underset{\sim}{\hat{w}} = \underset{\forall w_i}{\operatorname{argmin}} MSE = \underset{\forall w_i}{\operatorname{argmin}} E\left(\left(\sum_{\forall i} w_i X_i - \mu_i\right)^2\right)$ *such that,* $\sum_{i=1}^{z} w_i = 1$

Using the Lagrange multiplier approach,

$L = E\left(\left(\sum_{\forall i} w_i X_i - \mu_i\right)^2\right) + \lambda\left(\sum_{i=1}^{z} w_i - 1\right)$

The first order conditions with respect to the weights are as in the following for which there will be z such equations for each i

$$\frac{dL}{dw_j} = E\left(2\left(\sum_{\forall i} w_i X_i - \mu_i\right) X_i\right) + \lambda \tag{3.1}$$

This is then set to 0, which then results in w_i to be represented by the following :

$$\frac{dL}{d\lambda} = \sum_{i=1}^{z} w_i - 1 = 0 \tag{3.2}$$

from (1),

$\sum_{\forall i} w_i E\left(X_i X_j\right) - \mu_i^2 + \frac{\lambda}{2} = 0$

which is then equal to,

$\sum_{\forall i} w_i c_{ij} + \frac{\lambda}{2} = 0$

$\sum_{\forall i} \underset{\sim}{w} = -\frac{\lambda}{2}$

Therefore, $\underset{\sim}{w} \propto C^{-1}\underset{\sim}{1}$, where $\underset{\sim}{1}$ is a column of ones, thus giving a vector of weights for each expert.

Corollary 1 *If $\sigma_i^2 = \sigma_j^2\ \forall\ i, j$, and $\mu_i = \theta$, $\forall$, i, then $\underset{\sim}{w}$ is a function of Π.*

3.2.1 Examples demonstrating the theorem

Here, we provide two hypothetical examples demonstrating the between expert aggregation and within expert aggregation using Theorem 1. The example assumes a set of four experts each providing us their best estimates on four questions. Each of the question is considered independent of each other and each expert is assumed to be unbiased, which means that we assume that the experts are providing us the true value of the unknown parameter under consideration. The optimal weights are computed using Theorem 1 which minimizes the MSE. Example 3.1. Let us consider Table 3.2 that summarises the experts judgement on four questions:

Table 3.2: Example 3.1: Experts judgement on four questions

	E1	E2	E3	E4
Q1	0	0.5	0.3	0.08
Q2	0	0.5	0.6	0
Q3	0.2	0	0.2	0.5
Q4	0.99	0.35	0.1	0

We then compute the mean and covariances of these assessments across all questions. Therefore,

$\mu_1 = 0.220$, $\mu_2 = 0.275$, $\mu_3 = 0.225$, and $\mu_4 = 0.360$

The covariance is calculated using $c_{ijk} = Cov\left(X_{ki}, X_{kj}\right)$ is the covariance between expert i

and expert j's assessment over k^{th} question; where $Cov\left(X_{ki}, X_{kj}\right) = E\left(\left(X_{ki} - \mu_{ki}\right)\left(X_{kj} - \mu_{kj}\right)\right)$. Therefore the covariance matrix is given as:

Table 3.3: Covariance matrix of the assessments

	E1	E2	E3	E4
E1	0.2220	-0.0184	-0.0727	-0.0242
E2	-0.0184	0.0556	0.0266	-0.0519
E3	-0.0727	0.0267	0.0466	-0.0166
E4	-0.0242	-0.0519	-0.0166	0.0574

Then using $\rho_{ijk} = \frac{Cov\left(X_{ki}, X_{kj}\right)}{\sigma_{ki}\sigma_{kj}}$ is the correlation between expert i and expert j's assessment on question k we get the correlation matrix as Table 3.4

Table 3.4: Correlation matrix of the assessments

	[E1]	[E2]	[E3]	[E4]
[E1]	1	-0.1653	-0.7138	-0.2141
[E2]	-0.1653	1	0.5233	-0.9185
[E3]	-0.7138	0.5233	1	-0.3219
[E4]	-0.2141	-0.9185	-0.3219	1

Given the covariance matrix in Table 3.3, we can then use Theorem 1, which states that the weights that minimize MSE is the inverse of the covariance matrix multiplied by a vector of ones, therefore, we calculate the inverse of the covariance matrix C. Further, we can then compute the weights and normalise them so that the sum of the weights add to 1. Therefore, the weights are: $E1 = 0.0077$, $E2 = 0.7855$, $E3 = 0.1809$ and $E4 = 0.0257$. According to the weights received by the experts, Expert 2 has received the highest weight whereas Expert 1 has received the least weight. If we check the correlation matrix, Table 3.4, we see that Expert 1 and Expert 4 are negatively correlated with all other experts but Expert 4 is highly negative correlated as opposed to Expert 1 therefore Expert 4 receives a higher weight than Expert 1. Whereas Expert 2 is highly negatively correlated with Expert 4 and highly positively correlated with Expert 3, therefore receives a higher weight than all other experts. Expert 3 receives the second highest weight as it is highly negatively correlated with Expert 1. Thus this kind of aggregation takes into account the correlations that exists among the experts. It is hard to make a claim as to whether Expert 1 is a good expert or whether Expert 2,3 and 4 are enough for the assessment and Expert 1 is a redundant expert but this claim cannot be made because there is an underlying assumption that all experts are unbiased and they are giving true values of the unknown parameter. Though we cannot make any claims around which is the best expert out of the four experts however while aggregating it is worth taking the correlation of their assessments into account.

Example 3.2: In this example, we take into account the within expert correlation while aggregating. So we consider a different set of data for four experts on four independent questions.

Table 3.5: Example 2: Expert assessments on four independent questions

	Q1	Q2	Q3	Q4
E1	0.9722	0.6763	0.5806	0.3796
E2	0.0980	0.2474	0.6962	0.1571
E3	0.7718	0.7475	0.9909	0.3367
E4	0.6445	0.5815	0.6299	0.9982

Given these assessments, the means are then calculated across experts judgement, $\mu_1 = 0.62166$, $\mu_2 = 0.5632$, $\mu_3 = 0.7244$ and $\mu_4 = 0.4679$

Therefore, the covariance will be constructed using the column means.

Table 3.6: Covariance of the expert assessments

	E1	E2	E3	E4
E1	0.0608	-0.0184	0.03411	-0.0322
E2	-0.0184	0.0736	0.0500	-0.01763
E3	0.0341	0.0500	0.0744	-0.04631
E4	-0.0322	-0.0176	-0.0463	0.0367

Given the covariances and means, we could then calculate the correlations as in Table 3.7.

Table 3.7: Correlation between the experts assessment

	E1	E2	E3	E4
E1	1	-0.2756	0.5067	-0.6832
E2	-0.2756	1	0.6758	-0.3389
E3	0.5067	0.6758	1	-0.8854
E4	-0.6832	-0.3389	-0.8854	1

Given the covariance matrix in Table 3.6, we can then invert the matrix to get the optimal weights. Therefore the optimal weights obtained are as follows: $E1 = 0.2960$, $E2 = 0.1690$, $E3 = 0.0648$ and $E4 = 0.4700$. Here we see that Expert 4 has received highest weight and on assessing the correlation matrix in Table 3.7, we can see that Expert 4 is highly negatively correlated with all other experts. Expert 3 has received the lowest weight because this expert is positively correlated with Expert 2 and Expert 1 but highly negatively correlated with Expert 4. Again it cannot be concluded with certainty on which of these four experts could be possibly made redundant but given their correlations on their own assessments across questions, it is worth aggregating given the correlations.

3.3 Identifiability of Parameters

In statistics, identifiability of parameters is considered to make the inference possible, and it is therefore necessary to be able to mathematically establish these properties. As stated in section 2.3.2; non-parametric statistics lack asymptotic features and statistical power of tests for small sample sizes.

In the absence of correlations,

$E\left(X_{ki}X_{kj}\right) \rightarrow {\mu_n}^2 \,\forall\, i, j, k$, where i and j denote the experts, k is a counter for the number of questions and n represents the total number of questions. The total number of experts is denoted by $z \ni, i, j = 1, ..., z, \forall i \neq j$

Therefore,

$E\left(X_{ki}X_{kj}\right) = E\left(X_{ki}\right) E\left(X_{kj}\right)$, where $\forall i, j, k$.

The correlation coefficient ρ_{ijk} can be expressed in terms of the data as follows:

$$E\left({X^2}_{ki}\right) = {\sigma_{ki}}^2 + {\mu_{ki}}^2$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {x^2}_{ki}}{z\times n} \rightarrow \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {\sigma_{ki}}^2}{z\times n} + {\mu_{ki}}^2$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {x^2}_{ki}}{z\times n} - \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {\sigma_{ki}}^2}{z\times n} \rightarrow {\mu_{ki}}^2 \quad \forall\, i, j, k$$

$$E\left(X_{ki}X_{kj}\right) = c_{ijk} + {\mu_{ki}}^2$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} x_{ki}x_{kj}}{z\times n} = \rho_{ijk}\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z\times n} + {\mu_{ki}}^2$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \rho_{ijk}\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z\times n} \rightarrow {\mu_{ki}}^2 \ \forall\, i, j, k$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {x^2}_{ki}}{z\times n} - \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {\sigma_{ki}}^2}{z\times n} = \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \rho_{ijk}\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z\times n}$$

$$\rho_{ijk}\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z\times n} - \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {\sigma_{ki}}^2}{z\times n} = \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{z} x_{ki}x_{kj}}{z\times n} - \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {x^2}_{ki}}{z\times n}$$

$$\hat{\rho}_{ijk} = \frac{\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {x^2}_{ki}}{z\times n} + \frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} {\sigma_{ki}}^2}{z\times n}}{\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z\times n}}$$

$$\hat{\rho}_{ijk} = \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{x_{ki}x_{kj} - {x^2}_{ki} + {\sigma_{ki}}^2}{\sigma_{ki}\sigma_{kj}}\right)$$

The convergence is in probability and having expressed the correlation in terms of data, the mean can then be expressed in terms of data given dependence:

$$\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}x_{ki}x_{kj}}{z\times n} = \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{x_{ik}x_{jk}-x^2{}_{ki}+\sigma_{ki}{}^2}{\sigma_{ki}\sigma_{kj}}\right)\left(\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n}\right)+\mu_{ki}{}^2$$

$$\hat{\mu}_{ki}^2 = \frac{\sum_{i=1}^{z}\sum_{k=1}^{n}x_{ki}x_{kj}}{z\times n} - \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{x_{ik}x_{jk}-x^2{}_{ki}+\sigma_{ki}{}^2}{\sigma_{ki}\sigma_{kj}}\right)\left(\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n}\right)$$

$$\begin{aligned}\hat{\mu}_{ki}^2 &= \frac{1}{z\times n}\left(\sum_{i=1}^{z}\sum_{k=1}^{n}x_{ki}x_{kj} - \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{x_{ki}x_{kj}-x^2{}_{ki}+\sigma_{ki}{}^2}{\sigma_{ki}\sigma_{kj}}\right)(\sigma_{ki}\sigma_{kj})\right)\\ &= \frac{1}{z\times n}\left(\sum_{i=1}^{z}\sum_{k=1}^{n}x_{ki}x_{kj} - \sum_{i=1}^{z}\sum_{k=1}^{n}\left(x_{ki}x_{kj}-x^2{}_{ki}+\sigma_{ki}{}^2\right)\right)\\ &= \frac{1}{z\times n}\left(\sum_{i=1}^{z}\sum_{k=1}^{n}\left(x^2{}_{ki}-\sigma_{ki}{}^2\right)\right)\end{aligned}$$

Corollary 2 *MSE tends to* 0, *as* $z \to \infty$ *and* $n \to \infty$ *The asymptotic properties of the MSE can then be derived as:*

$$\begin{aligned}MSE\left(\hat{\mu_{np}}\right) &= E\left(\left(\sum_{\forall ki}w_{ki}X_{ki}-\mu_{ki}\right)^2\right)\\ &= E\left(X^2{}_{ki}\right)+\mu^2{}_{ki}-2\mu_{ki}E\left(X_{ik}\right)\\ &= \rho_{ijk}\frac{\sigma_{ki}\sigma_{kj}}{z\times n}\end{aligned}$$

$$\hat{\rho}_{ijk} = \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{x_{ki}x_{kj}-x^2{}_{ki}+\sigma_{ki}{}^2}{\sigma_{ki}\sigma_{kj}}\right)$$

$$MSE\left(\hat{\mu_{np}}\right) = \frac{1}{z\times n}\sum_{i=1}^{z}\sum_{k=1}^{n}\left(x_{ki}x_{kj}-x^2{}_{ki}+\sigma_{ki}{}^2\right)$$

$$\text{if } x_{ki}=0,\ MSE\left(\hat{\mu_{np}}\right) \to \sum_{i=1}^{z}\sum_{k=1}^{n}\left(\frac{\sigma_{ki}{}^2}{z\times n}\right)$$

Therefore, the MSE tends to 0, *as* $z \to \infty$ *and* $n \to \infty$.

3.4 Discussion

This chapter presented one mathematical method to aggregating correlated judgements under the assumptions of unbiased experts based on no assumptions on the parametric form of data. Although as the weights are directly proportional to the inverse of the covariance matrix, the assumption of a well defined positive definite covariance matrix is embedded.

There can be situations when the covariance matrix may not be positive definite; that can arise when the variances are zero. In such situations, a generalised inverse of the covariance matrix could be used instead for this proposed model to hold true. However, it is not always advisable to convert signular matrices into non-singular forms for modelling because the singularity could potentially lead to understanding the judgements that have been specified by the experts. Therefore, mathematically, this non-parametric model to aggregate correlated judgements would have a limited scope when the covariance matrices are singular.

It has been seen that the assignment of weights is of prime importance while aggregating judgements, especially when the judgements are correlated. It is well established in literature that if the experts are highly positively correlated then consulting multiple experts is same as consulting one expert because there is no gain in information; however, if the experts are negatively correlated, though there is more information gain around the parameter of interest, at the same time it also raises issues and concerns around the expert's understanding of the problem. It is fairly intuitive that in presence of any correlation whether negative or positive, the aggregated impact on the questions will be less than the arithmetic average when the judgements are treated independently. The method presented here assumed that the correlations are constant across the set of questions; as the questions are assumed to be independent. This is a fair assumption given that the questions are all related to assessing the same quantity of interest. When assessing the same quantity of interest, the shared knowledge shared work experience would play a crucial role in the assessments provided by the experts, hence assuming a constant correlation throughout their assessments is a fair assumption.

To conclude, correlations among expert judgements would invariably exist and this chapter provides a novel way of addressing the issue of aggregating judgements using these correlations in a sensible way while reducing the mean of the squared error using constrained optimisation with Lagrange multipliers.

Chapter 4

Empirical Bayes method for aggregation of expert judgement

4.1 Introduction

This chapter presents the empirical Bayes' method for aggregation of correlated expert's judgements. Section 4.2 introduces the normal/normal EB model for aggregation. This section then uses the exact same hypothetical examples from chapter three and allocates the weights to experts using EB method. Section 4.3 presents the identifiability of the parameters which further draws into identifying that the parameters can be estimated from data. Section 4.4 discusses and summarises the findings of this chapter.

4.2 Normal/Normal EB model for aggregation

The EB method for aggregation provides a novel way of aggregating the correlated experts judgements. It has already been established that the dependence between expert judgements should be modelled while aggregating the judgements otherwise the estimates tend to be erroneous (Chhibber and Apostolakis, 1993). While devising and dealing with an experimental design study, the dependence between and within the judgements are sensible ways for aggregation. The between and within expert variability has been discussed by Zio and Apostolakis (1997), but there is very little evidence in extant literature, which provides a suitable model to address the between and within expert variability. Although a lot of work has been done within the parametric framework of EB methodology, very little attempt has been made to structure a mathematical model around aggregation using correlated expert judgements. Thus, an attempt has been made to provide a modelling framework using EB methodology towards aggregating dependent expert judgements.

The biggest challenge that has been highlighted time and again is when to aggregate the probabilities elicited by the experts. The question that has intrigued many researchers is whether there is any impact of the new information on the aggregated posterior distribution, also known as 'external Bayesianity'. Bacco et. al, (2012) studied the impact of unanimity and compromise, which the decision maker and the experts together arrive at while combining probabilities. A two expert example using logitnormal distribution is proposed through their work where the potential source of information that is available to the expert and the statistician has been classified into three different categories. These categories are detailed information; specialist information available only to a specialist expert; and information concerning the decision problem at hand. In deciding on the most suited order of aggregation of the elicited probabilities, Bacco and colleagues (2012) concluded that coherent inference does not require experts to be combined via external Bayesian operator. It has been further stated that the disobedience of the EB operators takes place with regards to coherency because the information that is available to the experts is not necessarily available to the decision maker; therefore, which one of the information sources based on their categorisation is the expert extracting maximum information is unknown. Barker and Olaleyeln (2012) have also questioned the order of aggregation where the aggregation is based on quantiles and not probabilities, It has been found that combining experts earlier, before recomposition of the quantities, leads to smaller errors with less variance. Although the study is primarily based on simulations (the difference in errors is not huge) and it has been pointed out that in real world datasets, these differences might have a strong impact. Fairly recent mathematical aggregation literature showed that aggregating quantiles yielded a better aggregated result as opposed to aggregation of probabilities (Winkler et. al, 2013). However this thesis does not attempt to address the issues surrounding the aggregation order. This thesis provides mathematical frameworks to address dependencies and aggregate expert judgements when the judgements are expressed as point estimates. Further, the homogeneity or the extent to which the experts are similar in terms of the judgements that they are providing, that exists between the judgements through a repeated set of questions is an interesting aspect. EB approaches have been used to study this homogeneity, and it has been found that the inference accuracy increases when the pool of observations, and, as in this case, the expert's judgement is perfectly homogeneous (Quigley and Walls, 2011). As the relationship between the judgements provided by the experts is studied through correlations; the construction of the correlation matrix thus play a very important role in the aggregation analyses. However, linear correlation acts as a good measure when determining the co-movement given the Normal distribution scenario. It is well known that two variables can have a zero correlation but at the same time be strongly dependent (Ledoit and Wolf, 2003). Certainly in situations with non-normal distributions, linear correlation functions are able to conceal the strong co-dependence information that is contained in a full joint distribution. There could potentially be issues with the correlation matrix supporting negative eigenvalue and not being positive definite. The judgements could likely be negatively correlated on some questions and share a positive correlation on others. In situations when the true value is unknown to both the decision maker and the experts, then a strong positive correlation would give more confidence to the decision that would be concluded as an outcome of the aggregation (Jouini and Clemen, 1996). The EB method developed in this chapter provides a modelling framework that provides the inclusion of all possible (positive, negative and zero) correlations structure that might exist in a real scenario.

Consider the following notation:

Table 4.1: Notation used

z	represents the number of experts
n	represents the total number of questions
i, j	represent the experts where i and $j \in 1, 2, ..., z$

k	represents the question number; it acts as a counter for the number of questions, therefore $k \in 1, 2, ..., n$
X_{ki}	is the random variable representing expert i's assessment on question k
x_{ki}	is the realisation of X_{ki}
$\mu_{ki} = E\left(X_{ki}\right)$	is the expectation of expert i's assessment of question k
$\mu_i = E\left(X_{ki} \vert \mu_{ki}\right)$	is the conditional expectation of expert i's assessment of question k and is the true value for expert i's assessment $\forall$ k
$\sigma_{ki} = \sqrt{E\left(X_{ki} - \mu_{ki}\right)^2}$	is the standard deviation of expert i's assessment on question k
$c_{ijk} = Cov\left(X_{ki}, X_{kj}\right)$	is the covariance between expert i and expert j's assessment over k^{th} question; where $Cov\left(X_{ki}, X_{kj}\right)$ $= E\left(\left(X_{ki} - \mu_{ki}\right)\left(X_{kj} - \mu_{kj}\right)\right)$.
C_k	is the covariance matrix for k^{th} question where c_{ij} is the $(i, j)^{th}$ element $\forall$ k
C	represents the covariance matrix $\forall$ k
C^{-1}	represents the inverse of the covariance matrix $\forall$ k
$\rho_{ijk} = \dfrac{Cov\left(X_{ki}, X_{kj}\right)}{\sigma_{ki}\sigma_{kj}}$	is the correlation between expert i and expert j's assessment on question k

$\theta_k = E(\mu_{ki})$	is the prior mean of expert i's assessment of question k
$\tau_k = Var(\mu_{ki})$	is the prior variance of expert i's assessment of question k
Π	is the correlation matrix whose elements are σ_{ki}, c_{ij}, μ_k and ρ_{ij}
w_i	is the weight assigned to expert i
$\hat{\rho_{ijk}}$	is the estimator of ρ_{ijk}; $\hat{\rho_{ij}} = \frac{\sum\limits_{k=1}^{n} \hat{\rho_{ijk}}}{n}$
$\hat{w_i}$	is the estimator of w_i
$\hat{\theta_k}$	is the estimator of θ_k
$\hat{\tau_k}$	is the estimator of τ_k
$\hat{\mu_{ki}}$	is the estimator of μ_{ki}

The following set of general assumptions are made in the thesis on the proposed approaches:

$\rho_{ijk} = \rho_{ij} \ \forall \ k$

$Cov(X_{k_1 i}, X_{k_2 j}) = 0 \ \forall \ k_1 \neq k_2$

$\mu_{ki} = \mu_i \ \forall \ i$

μ_k	are assumed to be independent and identically distributed and follow $N(\theta_k, \tau_k)$
$\underline{X}_k$	follow multivariate Normal distribution $MVN\left(\underline{\mu}_k, C_k\right)$
$\hat{\mu}$	is the sample average

Given these set of assumptions, under the assumptions of Normality and Bayes rule, the parameters of the posterior distribution is given by Theorem 2; where $E\left(\mu_{ki}\right)=\theta_k$ which is the prior mean and $\tau_k^2=V\left(\mu_{ki}\right)$, which is the prior variance and these are calculated from the data. Also, $E\left(X_{ki}\right)=\theta$. However, the posterior parameters are specified for $k=1$ in theorem 2.

Theorem 2 *Given the prior distribution of n unbiased experts, where X_{ki} is the i^{th} expert's judgement on k^{th} question and this judgement is assumed to follow $N\left(\theta_k, \tau_k^2\right)$ and the likelihood follows a multivariate Normal distribution, $N\left(\mu_k, C_k\right)$. Then the posterior distribution is $N\left(\mu_1, \sigma_1^2\right)$ Normal with mean is a weighted average of the prior mean and the data and precision which is defined as the reciprocal of the variances (Bernado et. al, 2003), where*

$$\mu_1 = \theta\frac{\frac{1}{\tau_k^2}}{\left(\frac{1}{\tau_k^2}+\sum_{j=1}^{z}\sum_{i=1}^{z}c_{ij}\right)}+\frac{\sum_{i=1}^{z}c_{i.}x_{ki}}{\left(\frac{1}{\tau_k^2}+\sum_{j=1}^{z}\sum_{i=1}^{z}c_{ijk}\right)}; \textit{where}, c_{i.}=\sum_{j=1}^{z}c_{ijk}$$

$$\sigma_1^2=\sum_{i=1}^{z}\sum_{j>i,j\neq i}^{z}\left(\frac{1}{\sigma_{ki}^2(1-\rho_{ijk}^2)}-\frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)}+\frac{1}{\tau_k^2}\right)$$

and the weights are computed as

$$\hat{w}_i \propto C^{-1}+\frac{1}{\tau_k^2}$$

Proof 2 *In this derivation, the number of experts are denoted as z and the number of questions are denoted as n; consider a case where there are two expert assessments, x_{ki} and x_{kj}, a normal prior and a normal likelihood,*

$$\pi\left(\mu_i|\, x_{ki}\right)\propto\exp\left(-\frac{1}{2}\left(\frac{x_{ki}-\mu_i}{\sigma_{ki}}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{\mu_i-\theta}{\tau_k}\right)^2\right)$$

The posterior distribution can then be derived using the likelihood and the prior and it may *be expressed as:*

$$\pi\left(\mu_i \middle| x_{ki}, x_{kj}\right) \propto \exp\left(-\frac{1}{2\left(1-\rho_{ijk}^2\right)}\left[\begin{array}{c} \frac{(x_{ki}-\mu i)^2}{\sigma_{ki}^2} + \frac{\left(x_{kj}-\mu i\right)^2}{\sigma_{kj}^2} \\ -\frac{2\rho_{ijk}(x_{ki}-\mu_i)\left(x_{kj}-\mu_i\right)}{\sigma_{ki}\sigma_{kj}} \end{array}\right]\right)\exp\left(-\frac{1}{2}\left(\frac{\mu_i-\theta}{\tau_k}\right)^2\right)$$

On completing squares and collecting the terms for μ_i^2, μ_i and constants:

$$\frac{(x_{ki}-\mu_i)^2}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{(x_{kj}-\mu_i)^2}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} - \frac{2\rho_{ijk}\left(x_{ki}-\mu_i\right)\left(x_{kj}-\mu_i\right)}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} + \left(\frac{\mu_i-\theta}{\tau_k}\right)^2$$

$$= \mu_i^2\left(\begin{array}{c} \frac{1}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} \\ +\frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \end{array}\right) - 2\mu_i\left(\begin{array}{c} \frac{x_{ki}}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{x_{kj}}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} \\ +\frac{\theta}{\tau_k^2} - \frac{\rho_{ijk}\left(x_{ki}+x_{kj}\right)}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \end{array}\right) + const$$

where $const$ is the constant term comprising data.

$$= \left(\frac{1}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)}\right)$$
$$\left[\mu_i^2 - 2\mu_i\frac{\left(\frac{x_{ki}}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{x_{kj}}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{\theta}{\tau_k^2} - \frac{\rho_{ijk}\left(x_i+x_j\right)}{\sigma_i\sigma_j\left(1-\rho^2\right)}\right)}{\left(\frac{1}{\sigma_i^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)}\right)}\right] + const$$

Further simplification leads to the following, where the posterior distribution is expressed as a function of data (x_{ki}), μ_i, σ_{ki}^2 and the correlation coefficient, ρ_{ijk}.

$$= \left(\frac{1}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \right)$$

$$\left[\begin{array}{l} \mu_i^2 - 2\mu_i \frac{\left(\frac{x_{ki}}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{x_{kj}}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{\theta}{\tau_k^2} - \frac{\rho_{ijk}\left(x_{ki}+x_{kj}\right)}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \right)}{\left(\frac{1}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \right)} \\ + \left[\frac{\left(\frac{x_{ki}}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{x_{kj}}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{\theta}{\tau_k^2} - \frac{\rho_{ijk}\left(x_i+x_j\right)}{\sigma_i\sigma_j\left(1-\rho^2\right)} \right)}{\left(\frac{1}{\sigma_i^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)} \right)} \right]^2 \end{array} \right] + const$$

Therefore in the two expert case, it can be seen that μ_i, which is the mean of the expert assessments is expressed as a function of the prior mean, θ, the variance, σ_{ki}^2 and the prior variance, τ_k^2.

$$E\left[\mu_i \middle| x_{ki}, x_{kj}\right] = \frac{\frac{x_{ki}}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{x_{kj}}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{\theta}{\tau_k^2} - \frac{\rho\left(x_{ki}+x_{kj}\right)}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)}}{\frac{1}{\sigma_{ki}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)} + \frac{1}{\tau_k^2} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}\left(1-\rho_{ijk}^2\right)}}$$

$$= \frac{\sigma_{ki}^2\sigma_{kj}^2\tau_k^2\left(\frac{x_{ki}}{\sigma_{ki}^2} + \frac{x_{kj}}{\sigma_{kj}^2} + \frac{\theta\left(1-\rho_{ijk}^2\right)}{\tau_k^2} - \frac{\rho_{ijk}\left(x_{ki}+x_{kj}\right)}{\sigma_{ki}\sigma_{kj}} \right)}{\sigma_{kj}^2\tau_k^2 + \sigma_{ki}^2\tau_k^2 + \sigma_{ki}^2\sigma_{kj}^2\left(1-\rho_{ijk}^2\right) - 2\rho_{ijk}\sigma_{ki}\sigma_{kj}\tau_k^2}$$

$$= \left(\frac{1}{\begin{array}{l} \sigma_{kj}^2\tau_k^2 + \sigma_{ki}^2\tau_k^2 + \\ \sigma_{ki}^2\sigma_{kj}^2\left(1-\rho_{ijk}^2\right) - 2\rho_{ijk}\sigma_{ki}\sigma_{kj}\tau_k^2 \end{array}} \right) \left[\begin{array}{l} \left(\sigma_{kj}^2 - \rho_{ijk}\sigma_{ki}\sigma_{kj}\right)\tau_k^2 x_{ki} + \\ \left(\sigma_{kj}^2 - \rho_{ijk}\sigma_{ki}\sigma_{kj}\right)\tau_k^2 x_{kj} + \sigma_{ki}^2\sigma_{kj}^2\left(1-\rho_{ijk}^2\right)\theta \end{array} \right]$$

On generalising the two expert case onto four experts; the posterior distribution of μ_i is *normal with mean, μ_1,*

$$\mu_1 = \frac{\frac{x_{k1}}{\sigma_{k1}^2(1-\rho_{ijk}^2)} + \frac{x_{k2}}{\sigma_{k2}^2(1-\rho_{ijk}^2)} + \frac{x_{k3}}{\sigma_{k3}^2(1-\rho_{ijk}^2)} + \frac{x_{k4}}{\sigma_{k4}^2(1-\rho_{ijk}^2)} - \frac{\rho_{12k}(x_{k1}+x_{k2})}{\sigma_{k1}\sigma_{k2}(1-\rho_{ijk}^2)} - \frac{\rho_{13k}(x_{k1}+x_{k3})}{\sigma_{k1}\sigma_{k3}(1-\rho_{ijk}^2)} - \frac{\rho_{14k}(x_{k1}+x_{k4})}{\sigma_{k1}\sigma_{k4}(1-\rho_{ijk}^2)} - \frac{\rho_{23k}(x_{k2}+x_{k3})}{\sigma_{k2}\sigma_{k3}(1-\rho_{ijk}^2)} - \frac{\rho_{24k}(x_{k2}+x_{k4})}{\sigma_{k2}\sigma_{k4}(1-\rho_{ijk}^2)} - \frac{\rho_{34k}(x_{k3}+x_{k4})}{\sigma_{k3}\sigma_{k4}(1-\rho_{ijk}^2)} + \frac{\theta}{\tau_k^2}}{\frac{1}{\sigma_{k1}^2(1-\rho_{ijk}^2)} + \frac{1}{\sigma_{k2}^2(1-\rho_{ijk}^2)} + \frac{1}{\sigma_{k3}^2(1-\rho_{ijk}^2)} + \frac{1}{\sigma_{k4}^2(1-\rho_{ijk}^2)} - \frac{2\rho_{12k}}{\sigma_{k1}\sigma_{k2}(1-\rho_{ijk}^2)} - \frac{2\rho_{13k}}{\sigma_{k1}\sigma_{k3}(1-\rho_{ijk}^2)} - \frac{2\rho_{14k}}{\sigma_{k1}\sigma_{k4}(1-\rho_{ijk}^2)} - \frac{2\rho_{23k}}{\sigma_{k2}\sigma_{k3}(1-\rho_{ijk}^2)} - \frac{2\rho_{24k}}{\sigma_{k2}\sigma_{k4}(1-\rho_{ijk}^2)} - \frac{2\rho_{34k}}{\sigma_{k3}\sigma_{k4}(1-\rho_{ijk}^2)} + \frac{1}{\tau_k^2}}$$

Generalising the number of experts to z, the posterior mean can be written as,

$$\begin{aligned}
&= \sum_{i=1}^{z} \sum_{j>i, j\neq i}^{z} \left(\frac{\frac{x_{ki}}{\sigma_{ki}^2(1-\rho_{ijk}^2)} - \frac{\rho_{ijk}(x_{ki}+x_{kj})}{\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)} + \frac{\theta}{\tau_k^2}}{\frac{1}{\sigma_{ki}^2(1-\rho_{ijk}^2)} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)} + \frac{1}{\tau_k^2}} \right) \\
&= \sum_{i=1}^{z} \sum_{j>i, j\neq i}^{z} \left(\frac{x_{ki}\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)\tau_k^2 - \rho_{ijk}(x_{ki}+x_{kj})\sigma_{ki}^2(1-\rho_{ijk}^2)\tau_k^2 + \theta\sigma_{ki}^2(1-\rho_{ijk}^2)\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)}{\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)\tau_k^2 - 2\rho_{ijk}\sigma_{ki}^2(1-\rho_{ijk}^2)\tau_k^2 + \sigma_{ki}^2(1-\rho_{ijk}^2)\sigma_{ki}\sigma_{kj}(1-\rho_{ijk}^2)} \right) \\
&= \sum_{i=1}^{z} \sum_{j>i, j\neq i}^{z} \left(\frac{x_{ki}\sigma_{kj}\tau_k^2 - \rho_{ijk}(x_{ki}+x_{kj})\sigma_{ki}\tau_k^2 + \theta\sigma_{ki}^2\sigma_{kj}(1-\rho_{ijk}^2)}{\sigma_{kj}\tau_k^2 - 2\rho_{ijk}\sigma_{ki}\tau_k^2 + \sigma_{ki}^2(1-\rho_{ijk}^2)\sigma_{kj}} \right)
\end{aligned}$$

Therefore, the posterior mean is a weighted average of the prior mean and the data. Also, in order to estimate the parameters, the maximum likelihood estimation procedures have been followed, which results in the following estimates of mean and variance. In maximum likelihood estimation, the log of the likelihood function, in this case, the log of the multivariate normal likelihood function is maximized with respect to the parameters, θ and τ_k, respectively, and the first partial order derivative is equated to zero, the equation is then solved for the parameters. The second derivative is calculated to test whether the likelihood is maximized as it intended. Differentiating with respect to θ, and solving for θ gives,

$$\hat{\theta} = \sum_{i=1}^{z} \sum_{j>i, j\neq i}^{z} \frac{\left(1-\rho_{ijk}^{2}\right)\sigma^{2}{}_{ki}\sigma_{kj} + ki\tau_{k}^{2}\sigma_{kj} - \rho_{ijk}\left(x_{ki}+x_{kj}\right)\tau_{k}^{2}\sigma_{ki}}{\tau_{k}^{2}\sigma_{kj} + \sigma^{2}{}_{ki}\sigma_{kj}\left(1-\rho_{ijk}^{2}\right) - 2\rho_{ijk}\sigma_{ki}\tau_{k}^{2}}$$

The posterior mean can thus, be re-written as follows :

$$= \theta \frac{\frac{1}{\tau_k^2}}{\left(\frac{1}{\tau_k^2} + \sum_{j=1}^{z}\sum_{i=1}^{z} c_{ij}\right)} + \frac{\sum_{i=1}^{z} c_{i.}x_{ki}}{\left(\frac{1}{\tau_k^2} + \sum_{j=1}^{z}\sum_{i=1}^{z} c_{ijk}\right)}$$

$$where : c_{i.} = \sum_{j=1}^{z} c_{ijk}$$

$$E\left[\mu_i | x_{ki}\right] = w_0 E\left[\mu_i\right] + \sum_{i=1}^{z} w_i x_{ki}$$

$$w_i = \frac{c_{i.}}{\left(\frac{1}{\tau_k^2} + \sum_{j=1}^{z}\sum_{i=1}^{z} c_{ijk}\right)}$$

$$w_0 = 1 - \sum_{i=1}^{z} w_i$$

4.2.1 Examples demonstrating the theorem

Consider Table 3.2, where four experts provided their assessments on four questions. Based on these assessments, the first two moments, i.e., the mean and the standard deviation were calculated for each question. According to Theorem 2, the aggregated weight is sum of the prior precision (reciprocal of the variance) and the prior variance of the data. We calculate the prior variance given the data in Table 3.2 which are : $\sigma_1^2 = 0.22202500$, $\sigma_2^2 = 0.05562500$, $\sigma_3^2 = 0.02666667$ and $\sigma_4^2 = 0.05743333$. The prior precision is calculated from the data which is $\tau_1 = 0.2256841$, $\tau_2 = 0.3201562$, $\tau_3 = 0.2061553$ and $\tau_4 = 0.4450468$. Given the correlation between these assessments as in Table 3.4, the weights are then calculated as follows :

Table 4.3: Weights assigned to each expert for between expert correlation (based on Table 3.2.)

E1	0.2835
E2	0.2497
E3	0.3214
E4	0.1452

On comparison, Expert 2 and Expert 3 have also received higher weights as opposed to Expert 1 and Expert 4. Expert 1 received the least weight according to the EB method but the same method assigned Expert 4 the least weight. This is because of the prior variance of Expert 4. A higher prior variance leads to less homogeneity in the expert assessments while a low prior variance leads to a lower MSE (Quigley, et. al, 2011). On comparing the prior variance, it can be seen that Expert 4 has a higher prior variance of 0.4450468 as opposed to all other experts; therefore it is reinforcing the literature on EB method that prior variance has an impact on the aggregation given the correlations which is absent in the non-parametric method to aggregation.

Now considering Table 3.5, where four experts have provided their assessments on four independent questions and the covariance matrices calculated on Table 3.6, the weights assigned to each expert based on EB method is again computed as a sum of the prior variance of the data and the prior variance. Therefore, the weights assigned to experts are as follows:

Table 4.4: Weights assigned to each expert for within expert correlation (based on Table 3.5.)

E1	0.2469
E2	0.2186
E3	0.2115
E4	0.3227

The allocation of weights using the between experts and within expert aggregation is different. The between expert aggregation in Table 4.3 has assigned the highest weight to E3 whereas the within expert correlation, Table 4.4 has assigned highest weight to E4. The reason for the difference may be attributed to the high negative correlation between E3 and E4. It can also be seen that the allocation of weights for within expert correlation for both the methods i.e. the non-parametric and EB method is the same. Both the methods have assigned the highest weight to Expert 4 and the least weight to Expert 3. Expert 4 received the highest weight because the prior variance of Expert 4 is lowest $\tau_4 = 0.1916$ as opposed to the prior variances of Expert 1, $\tau_1 = 0.2465$, Expert 2, $\tau_2 = 0.2713$ and Expert 3, $\tau_3 = 0.2729$. While it can be easily seen that all other experts have received similar weights because their prior variances, in the examples, are similar as well. Though the correlations have played a critical role but as stated in the literature, the prior variances play an equal important role in assigning optimal weights to expert assessments given their dependencies.

4.3 Identifiability of Parameters

The asymptotic properties of EB estimates are investigated and it can be shown that $E_\theta\left[E\left(X_{ki}X_{kj}\right)|\theta_k\right] \to {\tau_k}^2 + {\mu_{ki}}^2\ \forall i,j,k$, when $\rho_{ijk} = 0$, i.e. in absence of any correlation, the $E_\theta\left[E\left(x_{ki}x_{kj}\right)|\theta_k\right] \to$ prior mean μ_{ki} and prior variance τ_k^2.

$$E_\theta\left[E\left(X_{ki}X_{kj}\right)|\theta_k\right] = \rho_{ijk}\sigma_{ki}\sigma_{kj} + E\left(X_{ki}\right)E\left(X_{kj}\right) = \rho_{ijk}\sigma_{ki}\sigma_{kj} + E\left({\theta^2}_k\right)$$

$$= \rho_{ijk}\sigma_{ki}\sigma_{kj} + Var(\theta_k) + \left(E\left(\theta_k\right)\right)^2$$

$$E_\theta\left[E\left(X_{ki}X_{kj}\right)|\theta_k\right] = \rho_{ijk}\sigma_{ki}\sigma_{kj} + {\tau_k}^2 + {\mu_{ki}}^2$$

$$\frac{\sum\limits_{i=1}^{z}\sum\limits_{j=1}^{z}\sum\limits_{k=1}^{n} X_{ki}X_{kj}}{z \times n} \to \rho_{ijk}\frac{\sum\limits_{i=1}^{z}\sum\limits_{k=1}^{n} \sigma_{ki}\sigma_{kj}}{z \times n} + {\tau_k}^2 + {\mu_{ki}}^2$$

Thus, the correlation can be expressed in terms of expert judgements and the uncertainties provided by the experts:

$$\frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} \to \rho_{ijk}\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n} + {\tau_k}^2 + {\mu_{ki}}^2$$

$$\frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \rho_{ijk}\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n} \to {\tau_k}^2 + {\mu_{ki}}^2$$

$$E_\theta\left[E\left(X_{ki}X_{kj}\right)|\theta_k\right] \to {\tau_k}^2 + {\mu_{ki}}^2$$

$$E_\theta\left[E\left({X^2}_{ki}\right)|\theta_k\right] = {\sigma_{ki}}^2 + {\tau_k}^2 + {\mu_{ki}}^2$$

$$\frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{z\times n} \to \frac{\sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n} + \left({\tau_k}^2 + {\mu_{ki}}^2\right)$$

$$E_\theta\left[E\left({x^2}_{ki}\right)|\theta_k\right] = \frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{z\times n} - \frac{\sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n} \to {\tau_k}^2 + {\mu_{ki}}^2$$

$$\frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \rho_{ijk}\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n} = \frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{z\times n} - \frac{\sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n}$$

$$\hat{\rho}_{ijk}\left(\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n}\right) = \frac{\sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n} + \frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{z\times n}$$

$$\hat{\rho}_{ijk} = \frac{\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n} + \frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{z\times n}}{\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n}}$$

$$\hat{\rho}_{ijk} = \frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj} + \sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2 - \sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}$$

Using the estimate for the correlation, the prior mean and variance can be expressed in terms of data:

$$\frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj}}{z\times n} - \left(\frac{\sum_{i=1}^{z}\sum_{j=1}^{z}\sum_{k=1}^{n} x_{ki}x_{kj} + \sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2 - \sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki}}{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}\right)\frac{\sum_{i=1}^{z}\sum_{k=1}^{n}\sigma_{ki}\sigma_{kj}}{z\times n} \to {\tau_k}^2 + {\mu_{ki}}^2$$

$$\frac{\sum_{i=1}^{z}\sum_{k=1}^{n} {x^2}_{ki} - \sum_{i=1}^{z}\sum_{k=1}^{n}{\sigma_{ki}}^2}{z\times n} \to {\tau_k}^2 + {\mu_{ki}}^2$$

The asymptotes suggest that the correlation coefficient can be entirely expressed in terms of the judgements and the standard deviations of these judgements. For a large sample size of expert judgements and questions, the covariance tends to the sum of the prior mean and the prior variance. The Mean Square Error (MSE) form is intractable, hence the impact of sample size on MSE is studied through simulations. However, when the experts provide same assessments for a particular question and they have the same uncertainty then, MSE tends to the prior variance:

$$MSE\,(\theta_k) \rightarrow {\tau_k}^2 \; \forall \, x_{ki} = x_{kj},\, \sigma_{ki} = \sigma_{kj}$$

$$\begin{aligned}
MSE\,(\theta_k) &= E\Big(\hat{\theta}_k - \theta_k\Big)^2 = Var(\hat{\theta}_k) + \Big(E\left(\hat{\theta}_k - \theta_k\right)\Big)^2 \\
&= E\left(\hat{\theta}_k^2\right) + E\left({\theta_k}^2\right) - 2E\left(\theta_k \hat{\theta}_k\right) \\
&= Var\left(\hat{\theta}_k\right) + \left(E\left(\hat{\theta}_k\right)\right)^2 + E\left({\theta_k}^2\right) - 2E\left(\theta_k \hat{\theta}_k\right) \\
&= Var\left(\hat{\theta}_k\right) \\
&= \frac{1}{\sum\limits_{i,j=1}^{z} \sum\limits_{k=1}^{n} \left(\frac{1}{\sigma^2{}_{ki}(1-{\rho_{ijk}}^2)} - \frac{2\rho_{ijk}}{\sigma_{ki}\sigma_{kj}(1-{\rho_{ijk}}^2)} + \frac{1}{{\tau_k}^2} \right)}
\end{aligned}$$

$$\hat{\rho}_{ijk} = \sum_{i=1}^{z} \sum_{k=1}^{n} \left(\frac{x_{ki}x_{kj} - x^2{}_{ik} + {\sigma_{ik}}^2}{\sigma_{ik}\sigma_{jk}} \right)$$

$$MSE\,(\theta_k) = \frac{1}{\sum\limits_{i,j=1}^{z} \sum\limits_{k=1}^{n} \left(\begin{array}{l} \dfrac{1}{\sigma^2{}_{ik}\left(1 - \left(\dfrac{x_{ki}x_{kj} - x^2{}_{ki} + {\sigma_{ki}}^2}{\sigma_{ki}\sigma_{kj}}\right)^2\right)} + \dfrac{1}{{\tau_k}^2} - \\ \dfrac{2}{\sigma^2{}_{ki}\sigma^2{}_{kj}\left(1 - \left(\dfrac{x_{ki}x_{kj} - x^2{}_{ki} + {\sigma_{ki}}^2}{\sigma_{ki}\sigma_{kj}}\right)^2\right)} \left(x_{ki}x_{kj} - x^2{}_{ki} + {\sigma_{ki}}^2\right) \end{array} \right)}$$

4.4 Discussion

The method proposed in this chapter is based on an empirical Bayes setting that uses the parametric normal distribution to model the prior and the likelihood. The structure of the aggregation within the Bayesian framework is cumbersome (Chhibber et. al, 1996). However, the significant use of this EB method permits the use of a broad correlation structure not necessarily restricting the correlations to be positive. Throughout the development of the approaches for aggregation, it has been assumed that the experts are well calibrated. The understanding of calibration while specifying probability values has been studied in both psychological literature as well as the statistical literature. Kahneman et. al, (1982) stated that in most real world decision problems, fairly large assessment errors make relatively little difference in the expected gain. However, this view has been argued by other scientists and it is stated that calibration is essential while assessments. Furthermore, it has been strongly suggested that any outcome that is achieved after a rigorous training, coherent subjectivists are well calibrated (Kahneman et. al, 1982). The calibration of expert judgements may also be viewed as a cognitive psychological process. It has been argued by Kahneman et. al, (1982), that people often tend to use simplification rules and heuristics to specify probability numbers. However, it has been counter argued that through proper training provided to the experts before the assessments are recorded, this issue of calibration may be tackled (Quigley and Walls, 2010). The posterior mean and variance within the EB framework are a combination of the weighted average of the expert judgements and the prior parameters. Based on the illustrated example it can be seen that the existence of less prior variation results in the experts getting more weight for their assessments; whereas if the assessments are heterogeneous then there is an increase in the uncertainty resulting in the experts receiving less weight for their assessments.

Chapter 5

Comparative study of the proposed methods

In this chapter, both the mathematical models namely, the non-parametric method and the EB method to aggregate correlated experts are compared on various criteria. These are namely, mathematical complexities, sensitivity to parameters, ease of understanding and convergence criteria. The sensitivity to parameters and the convergence of error (MSE) are examined through simulations. The simulations have been carried out in R software using standard R packages. Section 5.1 to 5.3 details the simulation outline and compares the results of both the models on the simulated data across various parameter settings. These in turn help in drawing conclusive remarks on the sensitivity of the models to the changes in parameters. Section 5.4 compares the ratio of MSE EB to MSE non-parametric, thus helping to draw a conclusion on the rate of convergence using both models. Section 5.5 studies the allocation of weight to experts given a diverse correlation structure. This section helps in formulating the conclusion around the allocation of weights using both the models.

5.1 Simulation of MSE using EB and non-parametric methods

In this section, structured simulations have been carried out for both the EB and non-parametric approaches to study the impact on MSE by varying and controlling for correlations and variances. The choice of the initial values of the parameters in terms of the number of experts and the number of questions is set to 2 so that a host of correlation matrices can be tested and the impact of the aggregation models developed in the thesis based across a broad range of correlations can be analysed. Literature does provide an evidence that there is a reduction in MSE with an increase in the number of experts or questions (Hogarth, 1987), hence the setting has been restricted to bivariate. As correlation and variations are the two most important concepts that form the focal point of this thesis, a range of values have been considered for both these parameters, i.e., C,ρ and τ. The initial values for the errors are set to 0 and 1000 simulation runs are been performed to test the models developed in the thesis. In order to study the impact of correlations on MSE, the following algorithmic steps have been followed:

1. Set simulation runs, $simruns$, to 1000
2. Set z, number of experts to 2
3. Vary τ^2 to $[0.001, 0.1, 0.5, 0.9]$
4. Vary ρ to $[-0.1, -0.5, 0.1, 0.5, 0.9]$

5. Generate θ using $rnorm$, to generate random normal variates in R

6. For each θ generate data, x, using $mvrnorm$ function in R

7. Calculate the sample mean from the data x

8. Calculate MSE EB as follows :

$$MSE\left(\theta\right)=\frac{1}{\sum_{i,j=1}^{z}\sum_{k=1}^{simruns}\left(\begin{array}{l}\frac{1}{\sigma^{2}{}_{ik}\left(1-\left(\frac{x_{ki}x_{kj}-x^{2}{}_{ki}+\sigma_{ki}{}^{2}}{\sigma_{ki}\sigma_{kj}}\right)^{2}\right)}+\frac{1}{\tau_{k}{}^{2}}-\\ \frac{2}{\sigma^{2}{}_{ki}\sigma^{2}{}_{kj}\left(1-\left(\frac{x_{ki}x_{kj}-x^{2}{}_{ki}+\sigma_{ki}{}^{2}}{\sigma_{ki}\sigma_{kj}}\right)^{2}\right)}\left(x_{ki}x_{kj}-x^{2}{}_{ki}+\sigma_{ki}{}^{2}\right)\end{array}\right)}$$

9. Calculate MSE non-parametric as follows:

   $MSE=E\left(\left(\sum_{i=1}^{z}w_{i}X_{i}-\mu\right)^{2}\right)$, where w_i can be computed from the covariance matrix generated by the simulated data. Detailed mathematics and derivations are in chapters three and four of the thesis.

10. Repeat 5 to 9 by varying values of τ and ρ

5.2 MSE EB

5.2.1 Impact of changing prior variance on MSE

Figure 5.1 shows that with a lower prior variance such as $\tau=0.0001$, the MSE EB approaches to 0 faster than with a higher prior variance. The four histograms demonstrate the different MSE EB distributions that occur in presence of different values of τ.

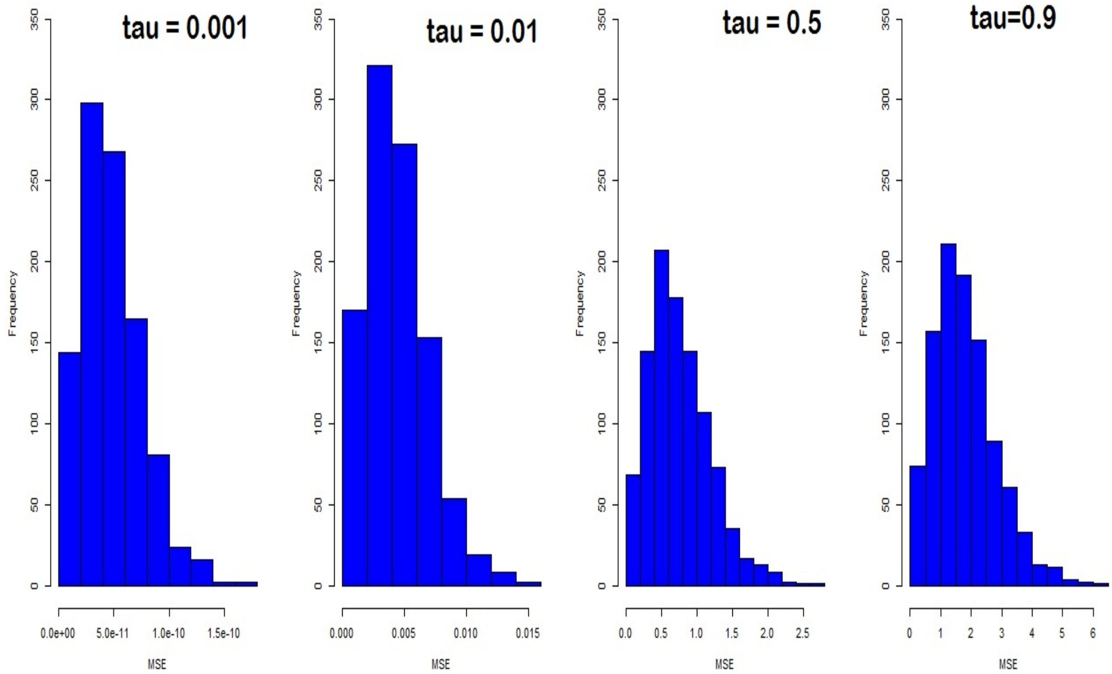


Figure 5.1: Figure showing the impact of changing prior variance on MSE EB

5.2.2 MSE EB : Impact of changing correlations on MSE

Figure 5.2 shows that with a negative correlation coefficient, the MSE EB approaches to 0 faster than with a positive correlation coefficient given all other parameters stay constant. For Figure 5.2, the prior variance i.e., τ, was fixed at 0.5 in order to assess the changes in MSE that is impacted solely by the correlation coefficient changes.

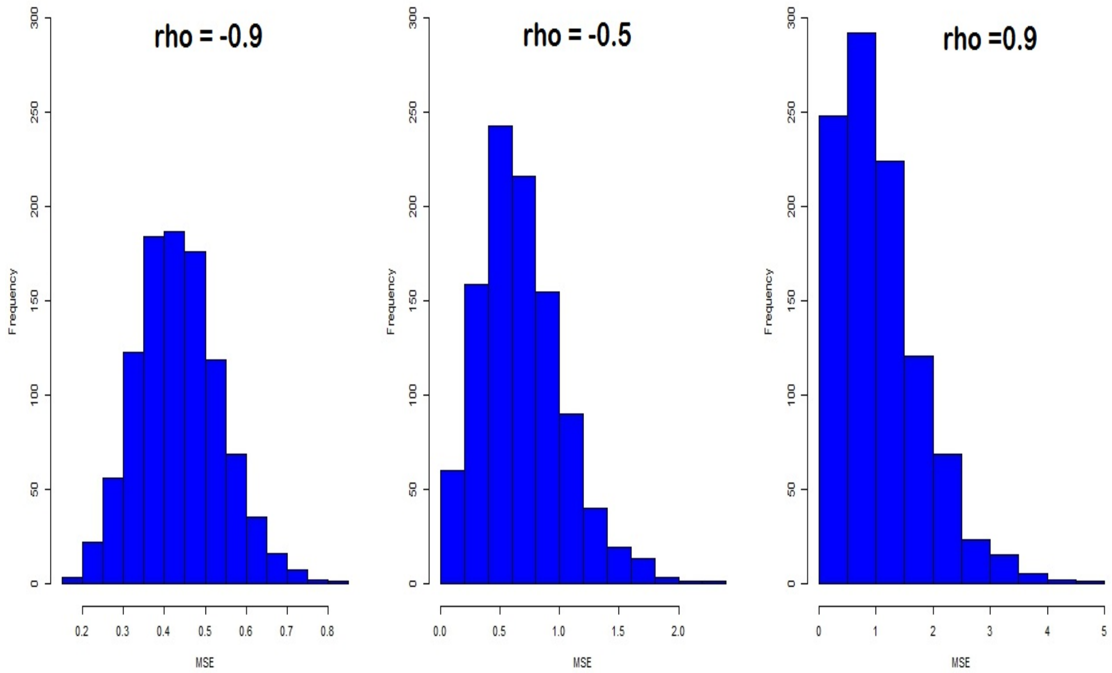


Figure 5.2: Figure showing the impact of changing correlation coefficient on MSE EB

5.3 MSE non-parametric

5.3.1 MSE non-parametric : Impact of changing prior mean on MSE

Figure 5.3 shows the impact of changing prior mean on MSE non parametric. A lower prior mean makes the MSE approach to 0 faster than a higher prior mean. However, it is worth noting that the MSE calculated using the non-parametric method in general leads to a slower rate of convergence of the MSE in comparison to the EB approach. An examination of the x-axis of Figure 5.3 shows a slower rate of convergence as opposed to the EB method.

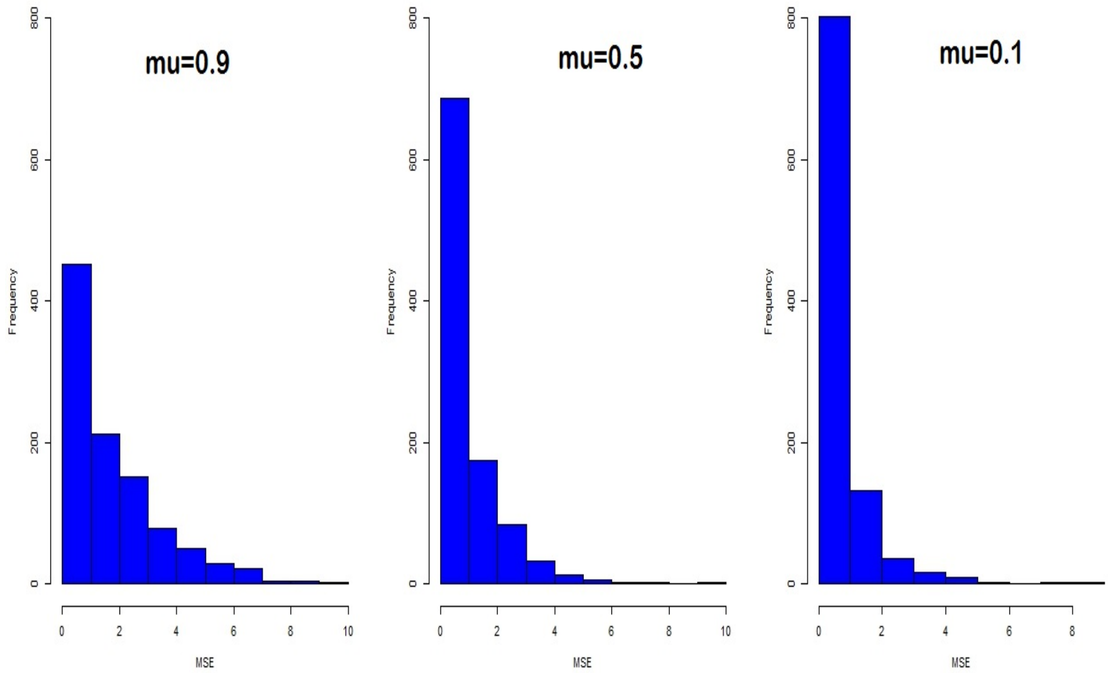


Figure 5.3: Figure showing the impact of changing prior mean μ on MSE non parametric

5.3.2 MSE non-parametric : Impact of changing correlations on MSE

Figure 5.4 shows the impact of changing correlation coefficient on MSE non parametric. A negative correlation makes the MSE approach to 0 faster than a positive correlation when all other parameters are kept constant.

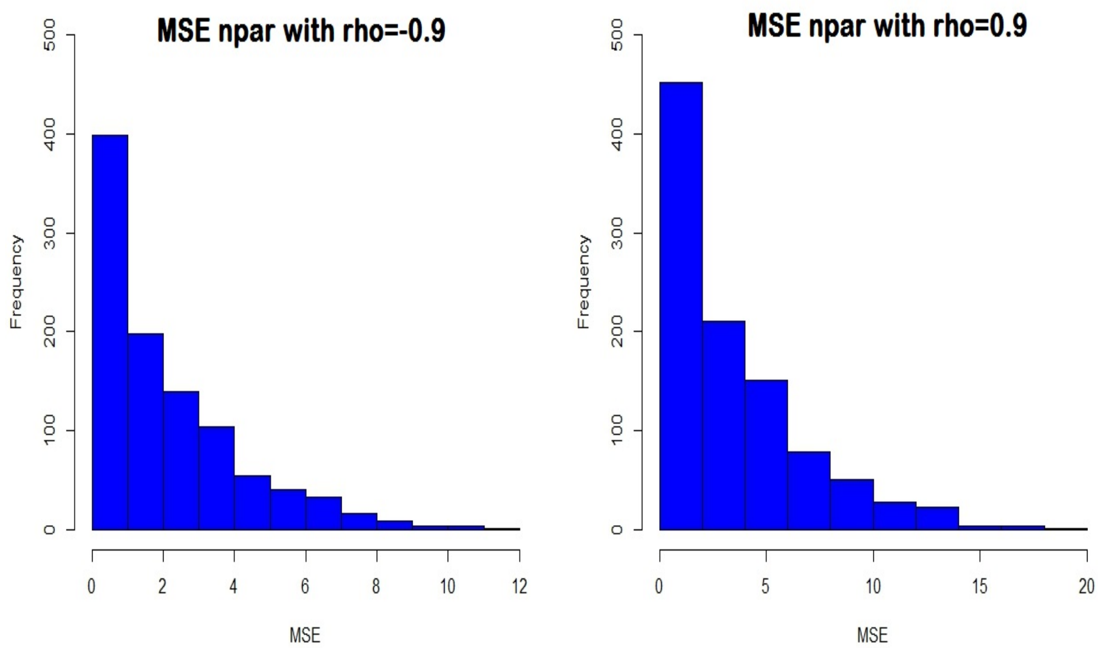
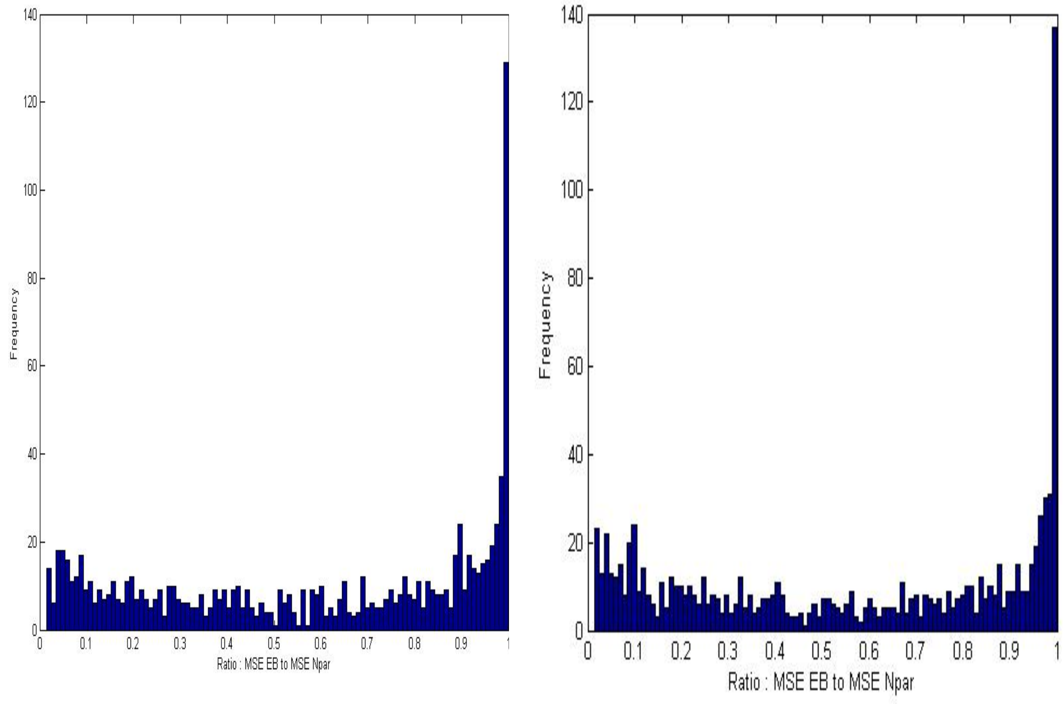


Figure 5.4: Figure showing the impact of changing correlation coefficient ρ on MSE non parametric

5.4 Comparison between MSE EB and MSE non parametric

It has been seen through the illustrated examples of chapter 3 and 4 that the allocation of weights to the experts in presence of correlation is similar. However, to arrive at a conclusive remark on the weight allocation using both the methods, weight allocation through different covariance structures are investigated in this section. Table 5.1 demonstrates a correlation structure among 3 experts where E1 and E2 are correlated with 0.5 whereas E1 and E3 are correlated with 0.9; the prior variance is fixed at 0.5 for all the three experts. Hence, using the prior variance and the correlation matrix, it can be seen that the non-parametric method has allocated least weights to E1 and E3 as opposed to E2. One possible reason for this difference is a lower correlation among E1 and E2 as opposed to the correlation between E1 and E3. Given a constant prior variance and an equal correlation of E2 with E1 and E3, the EB method has assigned a higher weight to E2 as opposed to others. In fact, both the methods have allocated the highest weight to E2; thus leading into a conclusion that lesser correlation leads to a higher weight allocation using both the methods.

Further investigations into the weighting process can be seen in Table 5.4 where the weights are equal given the independence in judgements of E1 and E3; and an assignment of a low weight to E2 due to his correlation with the other two experts. In Table 5.5, the assignment of weights is driven by a high prior variance of one expert, E2. Due to a high uncertainty in the judgements, E2 eventually receives the least weight as opposed to E1 and E3 whose prior variance is kept fixed at 0.1.

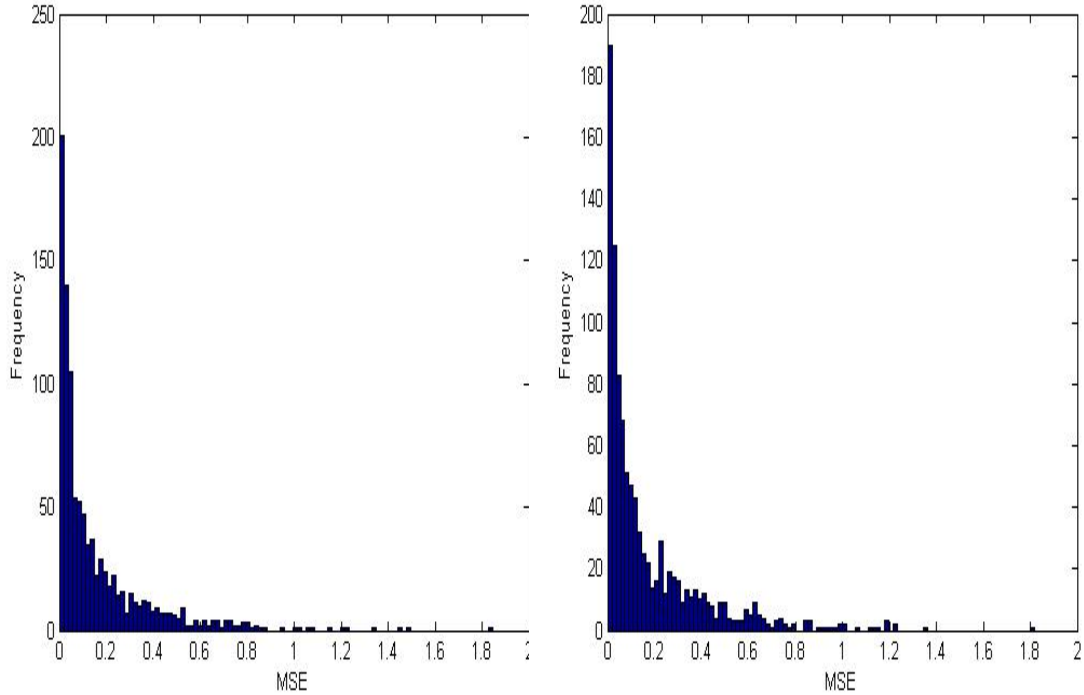


(a) MSE EB to MSE Npar when $\rho = 0.5$ (b) MSE EB to MSE Npar when $\rho = -0.5$

Figure 5.5: Figures 5.5(a) and 5.5(b) provide ratio of MSE EB to MSE Npar and it can be seen that the convergence for non-parametric is slower than EB.

5.5 Comparison of weights using EB and non-parametric approach

It has been seen through the illustrated examples of chapter 3 and 4 that the allocation of weights to the experts in presence of correlation is similar. However, to arrive at a conclusive remark on the weight allocation using both the methods, weight allocation through different covariance structures are investigated in this section. Table 5.1 demonstrates a correlation structure among 3 experts where E1 and E2 are correlated with 0.5 whereas E1 and E3 are



(a) MSE - EB

(b) MSE-non-parametric

Figure 5.6: The impact of an increase in the sample size of the experts will potentially lead to a faster convergence in MSE and this has been expressed through figures 5.6 (a) and 5.6(b). Impact of 5 experts where $\mu = 0.5$, $\tau = 0.1$, $\sigma = 0.1$, for all experts, however, $E1$ and $E3$ are correlated at 0.9 and all other experts are positively correlated at 0.5 on MSE. The simulation runs were 1000.

correlated with 0.9; the prior variance is fixed at 0.5 for all the three experts. Hence, using the prior variance and the correlation matrix, it can be seen that the non-parametric method has allocated least weights to E1 and E3 as opposed to E2. One possible reason for this difference is a lower correlation among E1 and E2 as opposed to the correlation between E1 and E3. Given a constant prior variance and an equal correlation of E2 with E1 and E3, the EB method has assigned a higher weight to E2 as opposed to others. In fact, both the methods have allocated the highest weight to E2; thus leading into a conclusion that lesser correlation leads to a higher weight allocation using both the methods.

Further investigations into the weighting process can be seen in Table 5.4 where the weights are equal given the independence in judgements of E1 and E3; and an assignment of a low weight to E2 due to his correlation with the other two experts. In Table 5.5, the assignment of weights is driven by a high prior variance of one expert, E2. Due to a high uncertainty in the judgements, E2 eventually receives the least weight as opposed to E1 and E3 whose prior variance is kept fixed at 0.1.

Table 5.1: Correlation matrix where E1 and E2 are correlated with 0.5 and E1 and E3 are correlated with 0.9

	E1	E2	E3
E1	1	0.5	0.9
E2	0.5	1	0.5
E3	0.9	0.5	1

Table 5.2: Weights assigned to experts

	non-parametric	EB
E1	0.2632	0.2911
E2	0.4737	0.4177
E3	0.2632	0.2911

Table 5.3: Correlation matrix where E1 and E2 are correlated with 0.5 and E1 and E3 are independent

	E1	E2	E3
E1	1	0.5	0
E2	0.5	1	0.5
E3	0	0.5	1

Given the correlation matrix in Table 5.1 and with $\sigma = 0.1$ for all the three experts, the weights then assigned are as follows:

Table 5.4: Weights assigned to experts

	non-parametric	EB
E1	0.5	0.4483
E2	0	0.1034
E3	0.5	0.4483

Table 5.5: Weights assigned to experts where ,τ for E2 $= 0.9$ and for E1,E3 are 0.1

	non-parametric	EB
E1	0.4554	0.4372
E2	0.0892	0.1256
E3	0.4554	0.4372

5.6 Discussion

In summary, this chapter provided a mathematical formulation that showed that the MSE using non-parametric method is slower in convergence to error than the MSE computed using EB method in presence of the correlated judgements. Further, it has been shown in this chapter that the weight allocation to experts using both the approaches is fairly similar; where a highly correlated expert is penalized and an independent expert is given higher weight to his/her assessments. This is what has been seen through the hypothetical examples of chapter three and four respectively. The MSE has a tractable feature while using the non-parametric method whereas using the EB approach, the form of MSE becomes cumbersome and does not have a closed form solution; thus highlights the mathematical complexity involving the EB method as opposed to the non-parametric method. Although it is established in the literature that a high sample size would lead to a lower MSE through the simulations reported in this chapter, it has been seen that the decrease in MSE not only depends on a large sample size but also on the correlations. The ratio of MSE EB to MSE non-parametric through the simulation studies have shown that EB performs better than method especially when the homogeneity of the experts is taken into consideration. For a higher prior variation, the MSE is large as opposed to a lower variation in the pool of experts who are chosen for the assessment task (Quigley et. al, 2011). However, the impact of this prior variation does not have any impact on the non-parametric method, thus making it evident that in absence of the true value of the quantities of interest, the non-parametric method is not inclusive of all available information about the experts. The results for weight allocations are consistent with the proposed literature surrounding dependencies because a sample of positively correlated experts is equivalent to having one expert. However, as these correlations cannot be ignored, when modelled they provide a meaningful way of allocating the weights to the judgements. The impact of a change in the sample size of experts and questions has been a subject of speculation for many different studies and through this study it has been reinforced that an increase in the number of questions or an increase in the number of experts does converge MSE faster to zero using both the methods. The theoretical derivation of the MSE for convergence have also been derived in chapters three and four respectively for the methods thus restating the impact of an increase in the sample size of experts and of questions.

Therefore, based on the study that has been conducted in this chapter, it may be concluded EB method is a preferred method for aggregation of correlated expert judgements as it is inclusive of the homogeneity of variance of the expert's judgements and has asymptotic features that exhibit a faster rate of convergence to error as opposed to non-parametric method.

Chapter 6

Application of proposed methods to FRB dataset

6.1 Introduction

This chapter provides the data analysis and aggregates the secondary dataset of expert judgements on the Forth Road Bridge, using the methods developed in this thesis. Section 6.2 provides a brief background to the case of the ageing bridge along with the data structure that is to be analysed. Section 6.3 presents an empirical data analysis of the expert judgement data and establishes the presence of correlation in the judgements. Section 6.4 is divided into subsections which allocate the weights to the experts based on the non-parametric and EB method developed in chapters three and four. Through the tabulated results in section (including subsections) 6.4, it can be seen that the weight allocation to experts given their correlation is fairly similar using both the methods. Section 6.5 discusses and summaries the findings of this chapter.

6.2 Background

Forth Road Bridge (FRB) is a suspension bridge near Edinburgh, Scotland which, was opened in 1964, and continues to serve as a vital link in Scotland's strategic road networks. The bridge has a main span of 1006 meters and side spans of 408 meters, (Colford and Clark, 2009). As a part of the Forth Estuary Transport Authority (FETA) requirement, the management and maintenance of the bridge is a carried out routinely. According to the results of the first internal inspection dated back in 2004/5, which were carried out by Faber Maunsell (Aecom) in association with Weidlinger Contractor C Spencer Ltd, it was concluded that there was an 8% strength loss in the main cables of the bridge (Colford and Clark, 2009). Along with this, it was also concluded that if the rate of deterioration could not be stopped or slowed down then the loading restrictions on the bridge may have to be put under serious considerations between 2014 and 2020. Following this, the second inspection was carried out in 2008 to benchmark the condition of the cables. The engineers concluded that the rate of deterioration was of the order of 10%, which basically showed an increase of 2% in the condition of the cables in past four years (Colford and Clark, 2009). The results from the internal inspections essentially stressed upon the fact that there could be absolutely no guarantee, given any further installations that were to be done to the bridge that would not prevent any further deterioration to the cables (Colford and Clark, 2009). Thus, following a cost and tendering task in August 2006, a feasibility study was commissioned to determine whether or not the main cables of the bridge could be either replaced or augmented (Andrew and Colford, 2006). One of the critical objectives of the study was to identify the most appropriate construction methods to be adopted should the main cables of the bridge be augmented or replaced . The complexity of the study increased as it involved weather conditions, transportations and the economy of surrounding areas along with the structural options. The feasibility study was scoped to include structural assessments of the existing and proposed cables; structural assessments of other aspects of the bridge such as decks, anchorages and saddles, among others; risk assessments, construction feasibility and safety assessments; traffic management; and economic assessments together with capital cost assessments. One of the quickest ways that would have dealt with the replacement of cables would have been to close the bridge for a period of three years, which in turn would have potentially eliminated the risks to the users of the bridge. However, it was considered politically unacceptable to close the bridge for this period (Colford et. al, 2009). A closer look at the scope of the feasibility study would have drawn the attention towards the condition of the anchorages, mainly because there was no access to the base of the rock tunnels, and the only inspection possible was to monitor for the movement at the strand anchor bearing plate. A feasibility study was conducted by Professors Quigley and Walls, Department of Management Science, University of Strathclyde to support W.A. Fairhurst and Partners, who were appointed as the consulting firm in statistically assessing the condition and capacity of the cable anchorages of the bridge (Quigley and Walls, 2010). The primary objective of the case study was to assess the long term ability to provide anchorage to the main cables with an adequate factor of safety.

In general, the conditions of transportation infrastructure has been of prime concern across the globe. The steep rise in population that adds onto the traffic volumes, limited funding available towards the maintenance of ageing bridges pushes the understanding and insights into the risk and reliability. Engineering tests such as direct pull-off test and others have been suggested as laboratory tests to compliment and contribute to the understanding of these conditions of such ageing bridges (Carrillo, 2012).

Along with the suggested and available laboratory methods for testing the conditions, expert judgements have been used with the framework of risk and reliability to tackle such engineering issues. The data for this thesis was available in terms of expert judgements to inform decisions concerning the selection of engineering tests to account for infrastructure such as for the anchorage capacity of the bridge. Expert judgement elicitation process was followed according to the Stanford Research Institutes general model for elicitation (Quigley and Walls, 2010) so as to reduce biases in the data collection process. Four experts were chosen based on their area of expertise, making sure that all relevant perspectives of the problem were covered. A Bayesian network was constructed and elicited to represent the expert judgements, which provides a model to support comparison of alternative engineering tests and sampling strategies. The elicitation process was twofold; one resulted in qualitative structuring of the problem and the other was quantification of the Bayesian Belief network (BBN) which was populated by the experts. Through the BBN, the conditional probabilities of the experts were collected. These conditional probabilities on the different states of the variables comprised the secondary dataset or data as referred throughout this thesis.

6.3 Empirical data analysis of the case study

The data consists of four expert judgements on variables such as Capacity, Tensioning North and South, Grout Condition North and South, Cracked Wires, Broken Wires, Surface Corrosion, Pitting, Capacity, Condition and engineering tests such as Direct pull of test, Lab wire inspection test, Lab tensile and On-site inspection test. Each expert provided his/her conditional probability estimate in terms of best guess. The data structure is similar to the Table 1.1 in chapter 1, where the best guess estimates are considered for several variables. A total of 1148 questions have been answered by 4 experts. To best understand the judgements that have been provided; an overview of the relationship between independent and dependent variables deduced from the Bayesian Belief Network that has been developed by Professors Quigley and Walls,(2010) is studied. The following relationships are deduced from the BBN

1. Bend, Grout condition North and South, Tensioning North and South are the independent variables
2. Pitting and Surface corrosion are dependent on the states of the Grout Condition
3. Cracked Wires are dependent on Pitting and Surface corrosion. Broken Wires are dependent on Cracked Wires
4. The variable Condition is dependent on Pitting, Surface Corrosion, Cracked Wires and Broken Wires
5. The engineering tests such as the Direct pull of test are dependent on Capacity, Bend and the Grout Condition

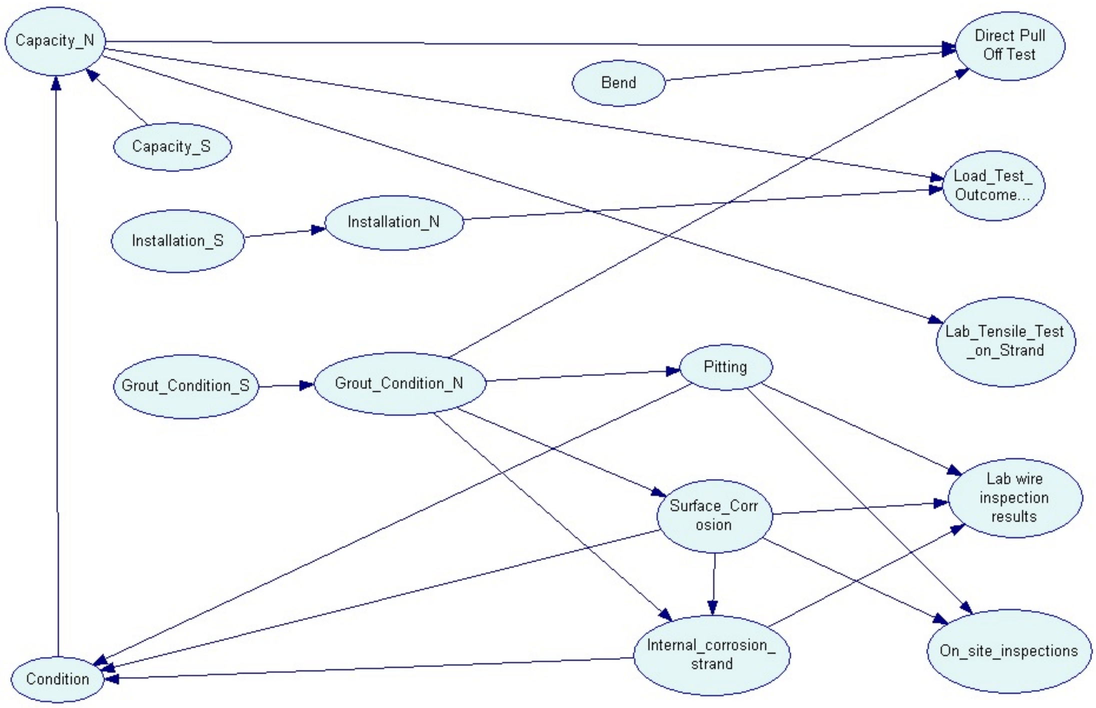


Figure 6.1: Bayesian Belief Network explaining relationships between variables, Quigley and Walls (2010)

6. The lab tensile strength test is dependent on Capacity and Tensioning

7. Lab Wire Inspection test is dependent on Pitting, Surface Corrosion, Cracked Wires and Broken Wires

8. Capacity is dependent on the variable Condition

9. On-site inspection test is dependent on Surface corrosion, Cracked Wires and Broken Wires.

Based on the secondary dataset of the Forth Road Bridge, five engineering tests were proposed. These were namely:

1. Direct Pull of test (DPT)

2. Lab wire inspection test (LWIT)

3. On site inspection test (ONIT)

4. Load test outcome (LTOS)

5. Lab Tensile test on strand (LTTS)

Each of these above mentioned tests had a number of outcomes. The experts have assessed these outcomes based on a combination of several conditions. For example, in order to understand the condition of the grout condition given that there is no pitting, the expert has provided his judgement on no movement which is an outcome of the direct pull-off test. Hence, if the grout condition is an independent event and pitting is another independent event, then the condition that there is no movement given that there is grout and given that there is no pitting gives a conditional probability that has been elicited from the experts. Each of these outcome combination for all the five tests are tabulated in Appendix A (A1 to A5). Further, these optimal test strategies have been studied using the BBNs and also using Bayes Linear methodology (Quigley, et. al, 2013; Quigley and Walls, 2010) for this dataset. Prior to using the mathematical models developed in the thesis, exploratory data analysis was carried out to provide an overview of the data. The presence of correlation has been assessed using the Pearson correlation coefficient which examines a linear relationship between any two variables. For example, Figure 6.2 shows the scatter plots of the best estimates of all experts across 1148 questions. It is seen that E2 and E4 are strongly correlated with a correlation coefficient of 0.72 as opposed to E1 and E2 who have a correlation of 0.29. In addition to the Pearson product moment correlation, Spearman's rho, which is the non-parametric measure of statistical dependence, is computed for all the four experts. The rho value for E2 and E4 is statistically significant at 0.63 indicates a high correlation between these two experts, indicating a strong correlation between E2 and E4.

Table 6.1: Product moment correlation between the experts

	E1	E2	E3	E4
E1	1			
E2	0.2893	1	0.5323	0.7213
E3	0.3478		1	0.6015
E4	0.4866			1

Table 6.2: Spearman's rank correlation between the experts

	E1	E2	E3	E4
E1	1	0.2249	0.2932	0.4528
E2	0.2249	1	0.4021	0.6333
E3	0.2932	0.4021	1	0.4923
E4	0.4528	0.6333	0.4923	1

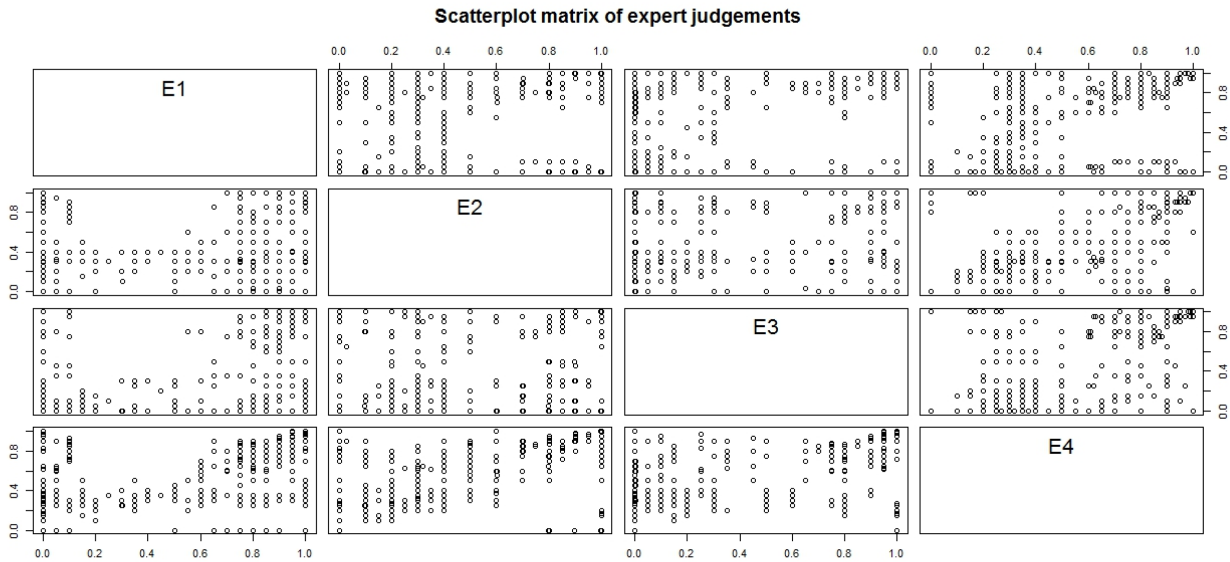


Figure 6.2: Scatterplot of expert judgements

To test for internal consistency as the same variable has been assessed by all the experts, the Cronbach's alpha measure has been calculated. The alpha value for the estimates is 0.641. The alpha value can also be interpreted as the correlation of an observed expert with all other experts. As a rule of thumb, alpha value for more that 0.6 is interpreted as a good and adequate measure of consistency (Cronbach, 1951). Therefore, according to the test results, it can be concluded that there is an adequate consistency across the dataset. Based on the analyses and results above, it can be concluded that the experts are correlated in their judgements.

6.4 Analyses using EB and non-parametric approaches

The problem of interest given this real dataset of correlated expert judgements is to understand what is the aggregated opinion on the set of proposed engineering tests? Having established a correlation structure overall between the expert judgements on the complete dataset of 1148 questions, then the dataset was broken into the five segments. Each segment belonged to each of the five proposed engineering test as enumerated above. The correlation matrix for each test across experts has been calculated and then both the proposed methods i.e. non-parametric and EB method have been applied to aggregate the judgements given the correlations for each test. The analysis section is broken into two sections; in section 6.4.1, the four expert judgements are aggregated using EB and non- parametric method across all questions and are compared to the simple arithmetic average. In section 6.4.2, the weight allocation to each expert using EB and non-parametric method is shown. Data from each of the engineering tests have been assumed to follow a normal distribution. The point estimates are then calculated using the EB and the non-parametric methods. Each of the estimate is tabulated for each test in Appendix A of this thesis.

6.4.1 Aggregation of judgements across questions using EB, non-parametric methods

Using the correlations, the EB point estimates and the non-parameteric estimates have been calculated and compared to simple airthmetic averages for all the engineering tests. It can be seen from Table A.1 to Table A.5 (attached to appendix A) that the EB estimates provide the lowest estimate for most questions in comparison to the non-parametric and simple arithmetic averages. The aggregated distribution computed using averages can be misleading as it ignores the presence of correlation between the expert judgements. The parametric EB method assumes a normal distribution which is symmetric in nature and incorporates a negative correlation matrix.

Aggregation methods on DPT

The correlation and covariance matrices across 96 questions have been calculated. The correlation matrix show a presence of negative correlation where E2 is negatively correlated with all the other three experts. Also, the prior variances for each expert are as follows : $E1 = 0.3038279$, $E2 = 0.4181079$, $E3 = 0.2889256$ and $E4 = 0.2995611$.

Table 6.3: Correlation Matrix - DPT Test

	E1	E2	E3	E4
E1	1	-0.1138209	0.4202440	0.8373401
E2	-0.1138209	1	-0.1392042	-0.1223879
E3	0.4202440	-0.1392042	1	0.3376372
E4	0.8373401	-0.1223879	0.3376372	1

Given the correlation matrix, the covariance matrix is calculated as follows:

Table 6.4: Covariance Matrix - DPT Test

	E1	E2	E3	E4
E1	0.09231140	-0.01445899	0.03689056	0.07621053
E2	-0.01445899	0.17481420	-0.01681615	-0.01532895
E3	0.03689056	-0.01681615	0.08347799	0.02922279
E4	0.07621053	-0.01532895	0.02922279	0.08973684

Given the variances and the covariance matrix, using the EB method, the weight allocated to each expert is as follows:

Table 6.5: Weight allocation EB method - DPT Test

	Weight allocated
E1	0.2480957
E2	0.1980516
E3	0.2844636
E4	0.2693891

The inverse of the covariance matrix is calculated in order to aggregate judgements using non-parametric method. The weights allocated by using the non-parametric method are as follows:

Table 6.6: Weight allocation non-parametric method - DPT Test

	Weight allocated
E1	0.05358375
E2	0.27847950
E3	0.37414749
E4	0.29378926

It can be seen from Table 6.5 and Table 6.6 that the experts have been allocated different weights based on the two different methods. The non-parametric method has penalised E1 because of his strong positive correlation with E4 whereas the EB method has penalised E2 for his negative correlation with all the other three experts. Interestingly enough, both E3 and E4 have received high weights through both the methods. E3 has received the highest weight using both the methods thus making E3 to be a highly reliable expert in this case.

Aggregation methods on LTOS

The correlation and covariance matrices across 16 questions have been calculated. The correlation matrix show a positive correlation amongst experts. Also, the prior variances for each expert are as follows : $E1 = 0.4949642$, $E2 = 0.3966815$, $E3 = 0.4745173$ and $E4 = 0.3901789$.

Table 6.7: Correlation Matrix - LTOS Test

	E1	E2	E3	E4
E1	1	0.04292017	0.6948907	0.7009184
E2	0.04292017	1	0.2340637	0.2760023
E3	0.69489065	0.23406367	1	0.6933693
E4	0.70091838	0.27600233	0.6933693	1

Given the correlation matrix, the covariance matrix is calculated as follows:

Table 6.8: Covariance Matrix - LTOS Test

	E1	E2	E3	E4
E1	0.244989583	0.008427083	0.16320833	0.13536458
E2	0.008427083	0.157356250	0.04405833	0.04271875
E3	0.163208333	0.044058333	0.22516667	0.12837500
E4	0.135364583	0.042718750	0.12837500	0.15223958

Given the variances and the covariance matrix, using the EB method, the weight allocated to each expert is as follows:

Table 6.9: Weight allocation EB method - LTOS Test

	Weight allocated
E1	0.2174177
E2	0.2961574
E3	0.2070601
E4	0.2793648

The inverse of the covariance matrix is calculated in order to aggregate judgements using non-parametric method. The weights allocated by using the non-parametric method are as follows:

Table 6.10: Weight allocation non-parametric method - LTOS Test

	Weight allocated
E1	0.199832666
E2	0.503325469
E3	0.007008922
E4	0.289832943

It can be seen from Table 6.9 and Table 6.10 that the experts have been allocated similar weights based on the two different methods. Both the methods have allocated E3 the least weight among all other experts. One possible reason could be that E3 is highly positively correlated with E1, E4 and also has a high prior variance as compared to all other experts. Whereas, even in spite of having a high variance, E1 has a slightly higher weight as opposed to E3 due to its lesser degree of correlation with other experts as opposed to E3. Both the methods have unanimously allocated the highest weight to E2 given its correlation with other experts.

Aggregation methods on LWIT

The correlation and covariance matrices across 416 questions have been calculated. The correlation matrix show a presence of negative correlation where E3 and E4 are negatively correlated. Also, the prior variances for each expert are as follows : $E1 = 0.2515237$, $E2 = 0.3058281$, $E3 = 0.3670020$ and $E4 = 0.2528545$.

Table 6.11: Correlation Matrix - LWIT Test

	E1	E2	E3	E4
E1	1	0.53650231	0.01692747	0.9353705
E2	0.53650231	1	0.02954798	0.5636581
E3	0.01692747	0.02954798	1	-0.0533432
E4	0.93537054	0.56365808	-0.05334320	1

Given the correlation matrix, the covariance matrix is calculated as follows:

Table 6.12: Covariance Matrix - LWIT Test

	E1	E2	E3	E4
E1	0.06326418	0.041269374	0.001562570	0.059488542
E2	0.04126937	0.093530816	0.003316452	0.043587685
E3	0.00156257	0.003316452	0.134690504	-0.004950149
E4	0.05948854	0.043587685	-0.004950149	0.063935405

Given the variances and the covariance matrix, using the EB method, the weight allocated to each expert is as follows:

Table 6.13: Weight allocation EB method - LWIT Test

	Weight allocated
E1	0.2727272
E2	0.2300197
E3	0.2028609
E4	0.2943922

The inverse of the covariance matrix is calculated in order to aggregate judgements using non-parametric method. The weights allocated by using the non-parametric method are as follows:

Table 6.14: Weight allocation non-parametric method - LWIT Test

	Weight allocated
E1	0.09473991
E2	0.16327965
E3	0.30296004
E4	0.43902039

It can be seen from Table 6.13 and Table 6.14 that the E4 has been allocated the highest weight using both the methods. While E4 is highly positively correlated with E1, he is also negatively correlated with E3. Interestingly, the EB method has assigned similar weights to E1 and E4 as they are highly positively correlated whereas the non-parametric method has penalised E1 and assigned E4 the highest weight given their correlation. One possible reason for this discrepancy in assignment of weight for E1 can be attributed to the prior variance of E1 and E4. As EB method takes into account the variances while allocating the weights, having similar sample variances does make E1 and E4 compatible in their judgements.

Aggregation methods on ONIT

The correlation and covariance matrices across 416 questions have been calculated. The correlation matrix show a presence of negative correlation where E2 is negatively correlated with all the other three experts. Also, the prior variances for each expert are as follows : $E1 = 0.2630612$, $E2 = 0.3200856$, $E3 = 0.3226339$ and $E4 = 0.2644492$.

Table 6.15: Correlation Matrix - ONIT Test

	E1	E2	E3	E4
E1	1	0.49510315	0.31809803	0.9379115
E2	0.4951032	1	0.04858244	0.4904952
E3	0.3180980	0.04858245	1	0.2588606
E4	0.9379115	0.49049522	0.25886061	1

Given the correlation matrix, the covariance matrix is calculated as follows:

Table 6.16: Covariance Matrix - ONIT Test

	E1	E2	E3	E4
E1	0.06920120	0.041688728	0.026997763	0.06524705
E2	0.04168873	0.102454792	0.005017131	0.04151864
E3	0.02699776	0.005017131	0.104092626	0.02208605
E4	0.06524705	0.041518638	0.022086054	0.06993335

Given the variances and the covariance matrix, using the EB method, the weight allocated to each expert is as follows:

Table 6.17: Weight allocation EB method - ONIT Test

	Weight allocated
E1	0.2571549
E2	0.2295041
E3	0.2325712
E4	0.2807698

The inverse of the covariance matrix is calculated in order to aggregate judgements using non-parametric method. The weights allocated by using the non-parametric method are as follows:

Table 6.18: Weight allocation non-parametric method - ONIT Test

	Weight allocated
E1	0.006165874
E2	0.273986858
E3	0.344075389
E4	0.375771879

It can be seen from Table 6.17 and Table 6.18 that E4 has been allocated the highest weight using both the methods. Due to a very high positive correlation between E1 and E4, the non-parametric method has penalised E1 but the EB method has given it the second highest weight due to its prior variance. E2 has received the lowest weight using both the methods.

Aggregation methods on LTTS

The correlation and covariance matrices across 8 questions have been calculated. The correlation matrix show a presence of negative correlation where E2 is negatively correlated with all the other three experts. Also, the prior variances for each expert are as follows : $E1 = 0.3537730$, $E2 = 0.4758451$, $E3 = 0.4350402$ and $E4 = 0.4862392$.

Table 6.19: Correlation Matrix - LTTS Test

	E1	E2	E3	E4
E1	1	0.6334924	0.5362010	0.3784896
E2	0.6334924	1	0.2035115	0.6421252
E3	0.5362010	0.2035115	1	0.7191765
E4	0.3784896	0.6421252	0.7191765	1

Given the correlation matrix, the covariance matrix is calculated as follows:

Table 6.20: Covariance Matrix - LTTS Test

	E1	E2	E3	E4
E1	0.12515536	0.10664286	0.08252429	0.06510714
E2	0.10664286	0.22642857	0.04212929	0.14857143
E3	0.08252429	0.04212929	0.18926000	0.15213000
E4	0.06510714	0.14857143	0.15213000	0.23642857

Given the variances and the covariance matrix, using the EB method, the weight allocated to each expert is as follows:

Table 6.21: Weight allocation EB method - LTTS Test

	Weight allocated
E1	0.3311689
E2	0.2149394
E3	0.2458123
E4	0.2080794

The inverse of the covariance matrix is calculated in order to aggregate judgements using non-parametric method. The weights allocated by using the non-parametric method are as follows:

Table 6.22: Weight allocation non-parametric method - LTTS Test

	Weight allocated
E1	0.57888202
E2	0.10879146
E3	0.22915698
E4	0.08316955

It can be seen from Table 6.21 and Table 6.22 that E1 has been allocated the highest weight whereas E4 has been allocated the least weight using both the methods. The allocation of weights to the experts is similar in this test for both the methods.

6.5 Allocation of weights - combined

In this section, the weight allocation across all the five proposed engineering tests is divided based on the method. Figure 6.3 shows the weight allocation for all four experts across the five tests using EB method and Figure 6.4 shows the weight allocation for all four experts across the five tests using the non-parametric method. It can be seen that the weight allocation has been fairly similar across tests using both the methods.

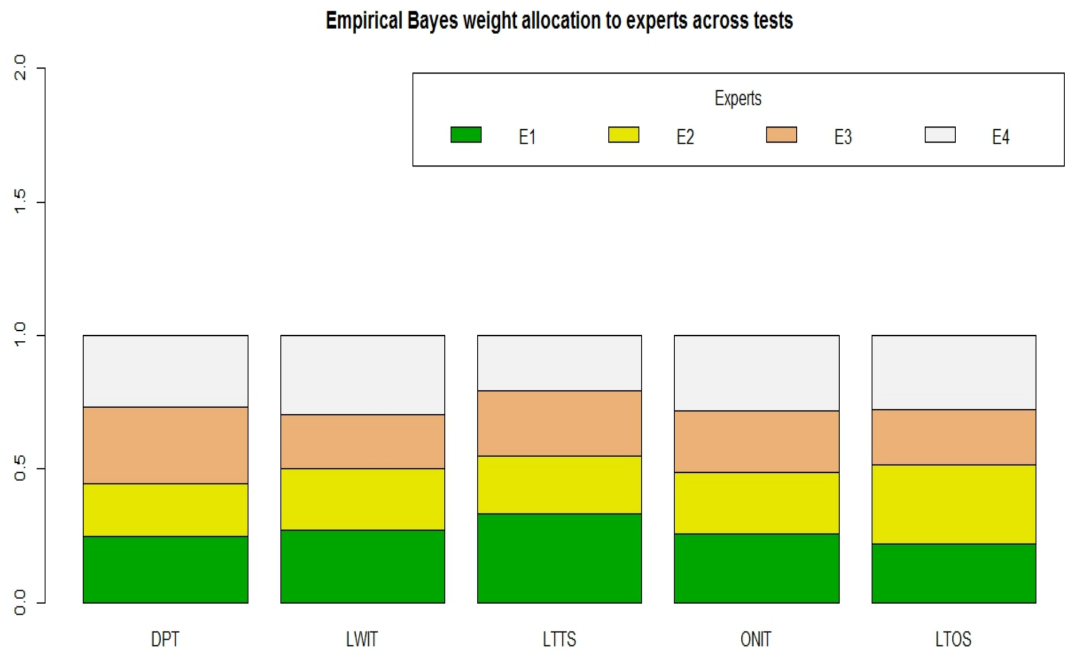


Figure 6.3: Figure6.3 depicts the EB assignment to the weights to different experts across different tests. It can be seen that Expert 1 has received the maximum weight in LTTS test while all the four experts have received similar weights in ONIT test. The experts have further received similar weights in LWIT test with E4 receiving a highest weight of 0.29 and E3 receiving a lowest weight of 0.20.

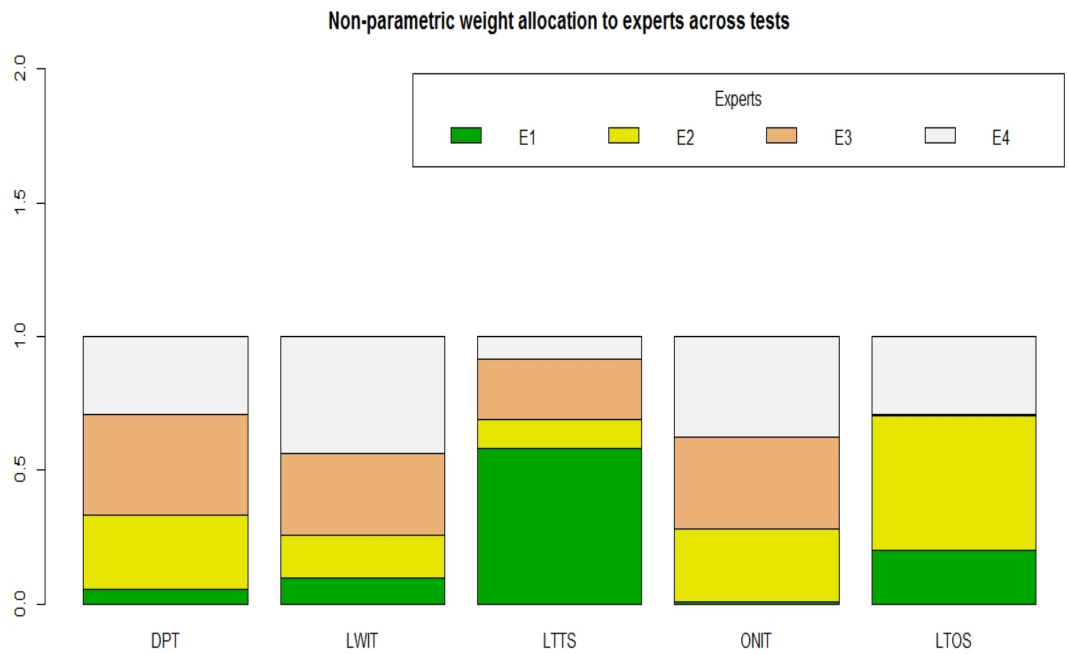


Figure 6.4: Figure6.4 depicts the assignment to the weights to different experts across different tests. It can be seen that Expert 1 has received the maximum weight in LTTS test and least weight in ONIT test and DPT test. E4 has received highest weight in LWIT test and ONIT test.

6.6 Discussion

The data used for this analysis is provided in appendix A (A1 to A5) of the thesis. It can be seen that the assignments of weights have been similar, where a positively correlated expert has received low weight as opposed to a negatively correlated or independent expert. Although in cases like in LWIT where the assignment of weight has been significantly different using both the methods; it can be seen that the use of prior variance in the EB method adds to the difference.

With the application of the proposed methods on the case study data, it has been revealed that both the methods penalize an expert for being highly positively correlated with one or more experts. However, it is worth mentioning that this aggregation is done in complete absence of true values of the underlying parameters. This further brings to question whether being positively correlated should really be penalized or whether being positive correlated is just a reflection of the experts' confidence in the true value of the underlying parameter of interest? Further, as it has been pointed out in the expert judgement literature that often positive correlation is required as it strengthens ones beliefs around the underlying parameter. This then brings to the question that whether having a set of all highly positively correlated experts is essential for a problem because it will be same as asking one expert rather than a whole set of expert? It is beyond the scope of this thesis to address the issue on an optimal number of experts for an aggregation problem but rather this thesis reinforces the knowledge that correlations among expert judgements is inevitable and there it must be accounted for rather than ignored while aggregating expert judgements.

The findings are summarized with respect to the conditions and the detailed comparison tables are attached to Appendix A (A1 to A5) of this thesis. It can be seen that some tests have a negative correlation matrix, which reflects some of disagreement that might exist between the experts across the different questions that they have assessed. On tracking back to the original dataset, there have been traces of difference in opinions that the experts have expressed during the elicitation process. These differences of opinions on assessing the states of different variables (also referred to as questions) impacts the correlation that exists between them, thus forcing some of the correlations in the judgements to be negative. It is also worth debating whether one method is better than the other and having referred back to the literature it can be seen that there is no predetermined mechanism by which a methodological superiority can be defined. Therefore, making an expert redundant because his / her judgements are identical to another may or may not be sensible because maybe the identical judgements do reflect the true answers. Also the real world data analyses showed that though the experts are identical on one test outcome, they differ on the others. Hypothetically, if the experts were similar on all questions and all test outcomes, then having two experts would be same as having one expert. Hence, from a pool of four experts making an expert redundant based on his/her performance on one test result would not be an ideal solution. The inference around which expert is the best or most reliable can also not be made because each expert has his/her own variance around the set of questions that they have defined. Hence, if the correlations of the experts together with their homogeneity in their variance can be best made use of while aggregating, informed decisions based on aggregation can be useful. The aim of this chapter was to apply both the models developed through chapter three and four to the secondary data of the Forth Road Bridge given the correlated judgements. It has been seen that the EB method made use of all available information including the prior variance of the experts while assigning fairly similar weights as the non-parametric method. Hence, based on the analyses of this dataset, EB method for weighting and aggregating is preferred over the non-parametric method.

Chapter 7

Discussion

This thesis highlighted the gap within the expert judgement literature where the dependencies between the judgements have not been accounted for while aggregating. This chapter provides the main findings of the individual chapters, a summary of the salient features of the mathematical approaches developed throughout the thesis, limitations and future research work while aggregating expert judgements when the judgements are correlated.

7.1 Summary of key findings

This thesis reinforces the view in the literature that correlations play a crucial role in the aggregation of expert judgements. Several reasons for the presence of this correlation have been touched upon in the literature review of chapter two of the thesis; like shared education shared knowledge among others. The aggregated judgement then provides a reflection of what the experts believe in, and also of the association and knowledge that they share of the problem at hand. A complete positive correlation among experts would be ideal and would render more confidence in the values of the parameters. However, the expert judgements could be negatively correlated (as is seen in chapter six of the thesis).

A structured method has been provided in this thesis towards weighting and aggregating judgements when the they are correlated. Also in cases when faced with a negative correlation between the judgements, it could be worth investigating or contacting the experts directly, which could also lead to potentially bringing in the behavioural aspects of the aggregation.

Both the parametric and non-parametric models developed in the thesis, have presented their pros and cons. Besides that, the novels help in providing the different forms of correlations. The results (attached to appendix A (A1:A5))show that the impact of aggregation in presence of any correlation is much lesser than simple arithmetic averages of the judgements. This holds true even when the experts are treated as independently (as seen in chapter five, Table 5.4). Both the EB and non-parametric methods used the first two moments and the covariance matrix to aggregate the judgements.

The non-parametric method is computationally less challenging as opposed to the EB approach, however, it lacks the asymptotic features of the EB approach. Also within the EB approach, the homogeneity, expressed in terms of the variance of the judgements provided, that exists among the pool of experts plays a crucial role in determining the overall aggregate variance as the combined variance is a combination of the prior and the likelihood variance. Hence, the more homogeneous the errors in the judgements are, the less uncertainty exists in the aggregate variance. This makes the EB method more appealing to aggregation than the non-parametric method in absence of the true value of the parameters of interest.

The weight allocation using both the EB and non-parametric method is similar. The weight allocation differs significantly from the equal weightings due to the presence of correlation that exists between the judgements. The EB method is mathematically cumbersome especially having no closed form solutions. This would be similar to the other Bayesian approaches that are used to aggregate expert judgements (Chhibber et. al, 1992) and as mentioned in chapter four, there are no closed form solutions to the squared loss of estimated errors. This is unlike the non-parametric approach, which provides a closed form solution to the estimated errors and is computationally less challenging for any sample of expert assessment.

The research into the methods has been conducted in multiple layers: a theoretical exposition of methods, an illustration of methods to simple hypothetical examples, examination of asymptotic properties of methods under a controlled simulation study, and application of methods to real expert judgement elicited in a secondary dataset.

7.2 Limitations and future work

Aggregation of expert judgements has been an important and productive field in terms of mathematical and as well as behavioural modelling. Aggregation of judgements has received significant importance in social policy making, environmental studies, epidemiological research, geological studies amongst many other varied disciplines (Bosetti et. al, 2012; Sabou et. al, 2013; Sol, 2013; Oz, and Mohamed, 2013). The mathematical constructs used across these studies vary from fuzzy logic to simple averages and it is primarily governed by the type of data. This area branches out into various other disciplines with a close connection to psychology and neuroscience that has been relatively less researched. The empiricism which is captured by the probabilities and assessed by the experts, supply singular vision in terms of correlation. This might exist and stress on the occurrence of the shared beliefs and knowledge which is a recurrent fear among the experts' judgement data.

There is a cost implication on acquiring and training experts for problems in real life. Through this work, it has been seen that correlations have a strong impact while aggregating opinion. It is a concern whether all positively correlated expert judgements are meaningless or whether that is an accurate reflection of the true value? Irrespective of the problem under review, it was beyond the scope of this thesis to study the optimality criteria surrounding the choice of experts.

However, the methods developed in this thesis can be improved with the use of different relevant prior distributions as opposed to the normal distribution. For example, the expert assessments could be assumed to follow a Beta prior or an uninformed prior and the likelihood could have been computed using MCMC algorithm for convergence. The non-parametric approach of minimising the MSE could be improved on by considering other loss functions. From the mathematical modelling perspective, the proposed assumptions surrounding the EB methods could be challenged and the sensitivity of a non-normal prior distribution and a different likelihood to that used in the thesis could be potentially used for aggregation (Frederick et. al, 2012). Although, the choice of the prior and likelihood is usually driven by the data.

Also, within the modelling approach, the non-parametric method proposed in this thesis depends on the use of Lagrange multipliers to achieve local minima for the MSE, given the constraint of weights. However, this method cannot be applied to error functions that are not convex in nature, as for the generalised Lagrange multiplier method to work, the underlying functions need to be either convex in nature or satisfy the Lipschitz continuity (Fernandez, 1997).

The dependencies could also be studied using Bayesian Belief Networks exploring external Bayesianity for multivariate expert judgements accounting for the overconfidence and pooling strategies for experts (Faria, 1996). Throughout this thesis, the correlations have been assessed using the product-moment correlations. However, if the rankings of the experts on the questions were available, then either the estimation of other correlations such as Kendall's tau or Spearman's rank correlation could be used to assess the dependence. With these measures, similar to Clemen and Reilly (1999)'s work, the likelihood function would then be expressed as a multivariate Gaussian copula instead of the normal/normal EB model as proposed in this thesis.

The potential impact of bias on aggregation has been theoretically presented in this thesis, attached to appendix B; however in cases where it is difficult to measure biases then the theoretical constructs around measuring biases would not have much impact on the mathematical models.

In situations, when the experts provide similar judgements across questions, and each of these experts also has the same variance across these repeated set of questions, this makes the covariance matrix singular. Although there are mathematical methodologies that deal with the problem of singular matrices, it is not advisable to convert every singular matrix into a non-singular form because the singularity does help in understanding the state of the underlying assessments. Hence, an amalgamation of behavioural methods and mathematical methods could potentially help deal with such situations. Singularity has not been an issue in this thesis but it could potentially have an impact with different real world datasets. There are also, several mathematical techniques in literature that deal with issues concerning singularity of the covariance matrix (McNeil, 1998; Jackal et al, 2000). Hence, if encountered with the problem of singularity the methods developed in this thesis can be used to aggregate given the techniques in literature.

On retrospection, the entire study on aggregation of correlated expert judgements could be potentially carried out adopting an experimental design pathway. Within this, a classical design of experiment method to the various tests and the judgements could be structured and the variability could potentially be explained in form of an analysis of variance (ANOVA) table summarising the variability between questions and also within experts. The potential challenges with this experimental design of study would be to test for its optimality. The problems with dealing with optimal designs within classical statistical literature are presented in Montgomery (2009). Whether a Bayesian experimental design study could be more suited to this context or a classical method is debatable as the expert data is subjective but because the treatment of the judgements is as data, it could be positioned within a classical design of experiment framework. Though an ANOVA method have stricter criteria of normality to understand and explain the underlying variation, but, it becomes interesting while formulating and extending the present research into answering the optimal criteria. This covers the number of experts and the number of questions with a trade off in between the experts' judgements variations. Adopting a design of experiment like approach to aggregation of experts' judgements could potentially lead to a controlled experimental method of expert judgement elicitation. However, structuring an experimental design often leads to understanding and studying cause and effect between factors and other variables. Therefore, in the context of experts' judgement aggregation, factors like cost and the uncertainty around the true value of the quantity of interest, could potentially be roadblocks in the experimental design pathway.

Another dimension that could potentially enhance the aggregation framework mathematically would be to incorporate the psychological factors that influence the correlations. The experts could also be given self-reported questionnaires on self-esteem and other factors (Rosenberg, M. (1965)). The data for analysis can then be a pooled data from the self-reported questionnaires and the expert judgements. Both the EB and non-parametric methods can then be applied to a pooled data which also comprises the psychological factors.

7.3 Summary of thesis contribution

Aggregation of judgements and opinions is incomplete without assessing the dependencies that exist between these judgements at different levels. Although the importance of the wisdom of the crowd has been highlighted in chapter one of this thesis, Eger (2013) presents several conditions under which such collective wisdom can completely fail. These conditions are based on a social learning process model which has applications across disciplines and most widely within social networks. The failure of the model raises issues around the rationality of the beliefs in the first place and the existence of bias in the judgements. The questions around the correctness of judgements that are being provided and also on how close to the true value these judgements are begged some vital questions in research.

However, in the absence of these true values, a rational and mathematically valid method towards the reduction of uncertainties could be seen as an alternative to providing an answer with fewer errors. Within the context of aggregation of judgements from multiple experts, this research makes the following contributions to knowledge:

1. Developed and evaluated mathematical methods that aggregate correlated expert's judgements.

2. Investigated and articulated the impact of correlations on aggregation of expert's judgements.

The thesis provides methodological contributions in form of two mathematical models; one based on an empirical Bayes method and the other based on non-parametric method, to aggregate dependent expert judgements data and through these methods, the effect and impact of correlations have been studied and investigated, re-emphasising on the importance of modelling the correlation. It has been seen that non-parametric methods lack the asymptotic features are opposed to EB methods, however, in presence of negative correlations, the MSE tends to converge to zero faster as oppose to in presence of positive correlation. While recent studies have argued the importance of weighting individual experts as opposed to equal weighting scheme (Bolger and Rowe, 2015), this research supports the literature claim of weighting individual experts. However, through this research it has been shown that the weighting done in presence of correlations and prior variances of the experts leads to an informed decision making under uncertainty. It has been seen that highly positively correlated experts receive lower weights for their assessments; in essence making one of the expert redundant, however, the question as to whether that particular assessment is the true value of the underlying parameter is debatable; especially when the true value of the underlying parameter of interest is not known. Based on the data analysis and simulations of this work, the EB method for aggregation is recommended for aggregation of correlated expert's judgements under uncertainty as opposed to the non-parametric method.

Bibliography

[1] Ambaum, Maarten HP. "Frequentist vs Bayesian statistics - a non-statisticians view." People (2012).

[2] Andrew, Andrew AS, and B. R. Colford. Forth Road Bridge-maintenance challenges. Fifth International Cable Supported Bridge Operators Conference. 2006

[3] Arbenz, Philipp, and Davide Canestraro. Estimating copulas for insurance from scarce observations, Expert Opinion and Prior Information: A Bayesian Approach. ASTIN Bulletin 42.1 (2012): 271-290.

[4] Armstrong, J. S. and W. School (1985). Long-range forecasting, Wiley New York etc.

[5] Baker, Erin, and Olaitan Olaleye. Combining Experts: Decomposition and Aggregation Order. Risk Analysis (2012).

[6] Bayarri, M. Jesus, and James O. Berger. The interplay of Bayesian and frequentist analysis. Statistical Science 19.1 (2004): 58-80.

[7] Ben-Arieh, David, and Zhifeng Chen. Linguistic-labels aggregation and consensus mea-

sure for autocratic decision making using group recommendations. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 36.3 (2006): 558-568.

[8] Bernardo, Jose M., ed. Bayesian Statistics Proceedings of the Seventh Valencia International Meeting. Oxford University Press, 2003

[9] Bolger, Fergus, and Gene Rowe. "The aggregation of expert judgment: do good things come to those who weight?." Risk Analysis 35.1 (2015): 5-11.

[10] Bonabeau, Eric. Don't trust your gut. Harvard Business Review 81.5 (2003): 116-123.

[11] Bonilla, Jesus Zamora. Optimal judgment aggregation. Philosophy of Science 74.5 (2007): 813-824.

[12] Bosetti, V., Catenacci, M., Fiorese, G., & Verdolini, E. (2012). The future prospect of PV and CSP solar technologies: An expert elicitation survey. Energy Policy.

[13] Broomell, Stephen B., and David V. Budescu. Why are experts correlated? Decomposing correlations between judges. Psychometrika 74.3 (2009): 531-553.

[14] Busemeyer, Jerome R., and James T. Townsend. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychological review 100.3 (1993): 432

[15] Budescu, David V., and Eva Chen. "Identifying expertise to extract the wisdom of crowds." Management Science 61.2 (2014): 267-280.

[16] Carlin, Bradley P., and Thomas A. Louis. Empirical Bayes: Past, present and future. Journal of the American Statistical Association 95.452 (2000): 1286-1289.

[17] Carrillo, Oscar Rafael Mata. Evaluating the bond durability of FRP-Concrete systems subjected to environmental exposures. Diss. Colorado State University, 2012

[18] Casella, George. An introduction to empirical Bayes data analysis. The American Statistician 39.2 (1985): 83-87.

[19] Chhibber, Sumeet, George Apostolakis, and David Okrent. A taxonomy of issues related to the use of expert judgments in probabilistic safety studies. Reliability Engineering & System Safety 38.1 (1992): 27-45.

[20] Chhibber, S., and G. Apostolakis. Some approximations useful to the use of dependent information sources. Reliability Engineering & System Safety 42.1 (1993): 67-86.

[21] Clemen, Robert T., and Robert L. Winkler. Combining probability distributions from experts in risk analysis. Risk analysis 19.2 (1999): 187-203.

[22] Clemen, Robert T., and Robert L. Winkler. Limits for the precision and value of information from dependent sources. Operations Research 33.2 (1985): 427-442.

[23] Clemen, Robert T., and Terence Reilly. Correlations and copulas for decision and risk analysis. Management Science 45.2 (1999): 208-224.

[24] Colford, B. R., and C. A. Clark. Feasibility study into the replacement or augmentation of the main cables of a long-span suspension bridge. Bridge Structures 5.2-3 (2009): 119-133.

[25] Colford, Barry R., and Colin A. Clark. Forth Road Bridge main cables: replace-

ment/augmentation study. Proceedings of the ICE-Bridge Engineering 163.2 (2010): 79-89.

[26] Cooke, Roger M. Experts in uncertainty: opinion and subjective probability in science. (1991).

[27] Cooke, Roger M. Problems with empirical Bayes. Risk Analysis 6.3 (1986): 269-272.

[28] Cox, David Roxbee, and David Victor Hinkley. Theoretical statistics. CRC Press, 1979

[29] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika 16(3): 297-334.

[30] Dayan, Peter, Laurence F. Abbott, and L. Abbott. Theoretical neuroscience: Computational and mathematical modeling of neural systems. (2001).

[31] Deely, J. J., and D. V. Lindley. Bayes empirical bayes. Journal of the American Statistical Association 76.376 (1981): 833-841.

[32] Devine, Owen J., Thomas A. Louis, and M. Elizabeth Halloran. Empirical Bayes methods for stabilizing incidence rates before mapping. Epidemiology 5.6 (1994): 622-630.

[33] Dutfoy, A. and R. Lebrun (2009). Dependence modelling with copula in probabilistic studies, a practical approach based on numerical experiments.

[34] Efron, Bradley. Bayesians, frequentists, and scientists. Journal of the American Statistical Association 100.469 (2005).

[35] Efron, Bradley, and Robert Tibshirani. Empirical Bayes methods and false discovery rates for microarrays. Genetic epidemiology 23.1 (2002): 70-86.

[36] Eger, Steffen. (Failure of the) Wisdom of the crowds in an endogenous opinion dynamics model with multiply biased agents. arXiv preprint arXiv:1309.3660 (2013).

[37] Faria, Alvaro Eduardo. Graphical Bayesian models in multivariate expert judgements and conditional external Bayesianity. Diss. University of Warwick, (1996).

[38] Fernandez, Luis A. Classroom note: On the limits of the Lagrange multiplier rule. SIAM review 39.2 (1997): 292-297.

[39] Fishburn, Peter C. Utility theory for decision making. No. RAC-R-105. Research Analysis Corp Mclean VA, 1970

[40] Flores, Benito E., and Edna M. White. Subjective versus objective combining of forecasts: an experiment. Journal of Forecasting 8.3 (1989): 331-341.

[41] Frederic, Patrizio, Mario Di Bacco, and Frank Lad. Combining expert probabilities using the product of odds. Theory and decision 73.4 (2012): 605-619.

[42] French, Simon. Decision theory: an introduction to the mathematics of rationality. Halsted Press, 1986

[43] Galton, Francis. Inquiries into human faculty and its development. Macmillan and Company, 1883

[44] Garthwaite, Paul, Ian T. Jolliffe, and Byron Jones. Statistical inference. Oxford University Press, 2002

[45] Gelman, Andrew. Exploratory data analysis for complex models. Journal of Computational and Graphical Statistics 13.4 (2004).

[46] Gelman, Andrew. Objections to Bayesian statistics. Bayesian Analysis 3.3 (2008): 445-449.

[47] Genest, Christian, Kevin J. McConway, and Mark J. Schervish. Characterization of externally Bayesian pooling operators. The Annals of Statistics 14.2 (1986): 487-501.

[48] Genest, Christian. Vincentization revisited. The Annals of Statistics 20.2 (1992): 1137-1142.

[49] Ghauri, Pervez. Designing and conducting case studies in international business research. Handbook of qualitative research methods for international business (2004): 109-24.

[50] Ghosh, Jayanta K., and R. V. Ramamoorthi. Bayesian nonparametrics. Springer, 2003

[51] Gigone, Daniel, and Reid Hastie. Proper analysis of the accuracy of group judgments. Psychological Bulletin 121.1 (1997): 149

[52] Girotto, Vittorio, and Michel Gonzalez. Solving probabilistic and statistical problems: A matter of information structure and question form. Cognition 78.3 (2001): 247-276.

[53] Gokhale, D. V., and S. James Press. Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. Journal of the Royal Statistical Society. Series A (General) (1982): 237-249.

[54] Goldstein, William M., and Robin M. Hogarth, eds. Research on judgment and decision making: Currents, connections, and controversies. Cambridge University Press, 1997

[55] Gong, Gail, Nathan Hannon, and Alice S. Whittemore. Estimating gene penetrance from family data. Genetic epidemiology 34.4 (2010): 373-381.

[56] Gu, Yuhong. A comparison of aggregation methods of subjective probability distributions. PhD diss., Fordham University, 2009

[57] Hajek, Alan, and Stephan Hartmann. Bayesian epistemology. 2001) A Companion to Epistemology, Wiley. ISBN 1405139005 (2010).

[58] Hammitt, James K., and Alexander I. Shlyakhter. The expected value of information and the probability of surprise. Risk Analysis 19.1 (1999): 135-152.

[59] Hammitt, James K., and Yifan Zhang. Combining Experts Judgments: Comparison of Algorithmic Methods Using Synthetic Data. Risk Analysis 33.1 (2013): 109-120.

[60] Hanea, A. M., et al. "I nvestigate D iscuss E stimate A ggregate for structured expert judgement." International Journal of Forecasting 33.1 (2017): 267-279.

[61] Hjort, Nils Lid, ed. Bayesian nonparametrics. No. 28 Cambridge University Press, 2010

[62] Hogarth, Robin M. A note on aggregating opinions. Organizational Behavior and Human Performance 21.1 (1978): 40-46.

[63] Hogarth, Robin M. Judgement and choice: The psychology of decision . John Wiley & Sons, 1987

[64] Hogarth, Robin M. Educating intuition. University of Chicago Press, 2001

[65] Howard, Ronald A. Decision analysis: practice and promise. Management science 34.6 (1988): 679-695.

[66] Jackel, Peter. Monte Carlo methods in finance. Stochastic Dynamics 3 (2001): 3-2.

[67] Jara, Alejandro, et al. DPpackage: Bayesian non-and semi-parametric modelling in R. Journal of statistical software 40.5 (2011): 1

[68] Jou, Jerwen. Multiple number and letter comparison: Directionality and accessibility in numeric and alphabetic memories. The American journal of psychology (2003).

[69] Jouini, Mohamed N., and Robert T. Clemen. Copula models for aggregating expert opinions. Operations Research 44.3 (1996): 444-457.

[70] Justin W. Eggstaff, Thomas A. Mazzuchi, Shahram Sarkani, The effect of the number of seed variables on the performance of Cookes classical model, Reliability Engineering & System Safety, Volume 121, January 2014, Pages 72-82, ISSN 0951-8320.

[71] Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. Judgment under uncertainty: Heuristics and biases. Cambridge University Press, 1982

[72] Kallen, M. J., and R. M. Cooke. Expert aggregation with dependence. Probabilistic Safety Assessment and Management. Elsevier Science, 2002

[73] Kass, Robert E. Statistical inference: The big picture. Statistical science: a review journal of the Institute of Mathematical Statistics 26.1 (2011): 1

[74] Kauko, Tom. Residential property value and locational externalities: On the complementarity and substitutability of approaches. Journal of Property Investment & Finance 21.3 (2003): 250-270.

[75] Kiefer, Nicholas M. Default estimation for low-default portfolios. Journal of Empirical Finance 16.1 (2009): 164-173.

[76] Lambrigger, Dominik D., Pavel V. Shevchenko, and Mario V. WÃijthrich. The quantification of operational risk using internal data, relevant external data and expert opinions. Journal of Operational Risk 2.3 (2007): 3-27.

[77] Ledoit, Olivier, and Michael Wolf. Honey, I shrunk the sample covariance matrix. UPF Economics and Business Working Paper 691 (2003).

[78] Lee, Yuh-Shiow. Effects of learning contexts on implicit and explicit learning. Memory & cognition 23.6 (1995): 723-734.

[79] Lichtendahl, Kenneth C., Yael Grushka Cockayne, and Robert L. Winkler. Is It Better to Average Probabilities or Quantiles?. Management Science (2013).

[80] Lin, Shi-Woei. Jackknife evaluation of uncertainty judgments aggregated by the Kullback Leibler distance. Applied Mathematics and Computation 218.2 (2011): 469-479.

[81] Lin, Shi-Woei, and Chih-Hsing Cheng. The reliability of aggregated probability judgments obtained through Cooke's classical model. Journal of Modelling in Management 4.2 (2009): 149-161.

[82] Lindley, Dennis V. The philosophy of statistics. Journal of the Royal Statistical Society: Series D (The Statistician) 49.3 (2000): 293-337.

[83] Lindley, Dennis V. Another look at an axiomatic approach to expert resolution. Management science 32.3 (1986): 303-306.

[84] List, Christian, and Clemens Puppe. Judgment aggregation: a survey. Working paper,

London School of Economics (forthcoming in: Anand P, Pattanaik P, Puppe C (eds) The handbook of rational and social choice. Oxford University Press), 2007

[85] Lyon, Aidan, and Eric Pacuit. The Wisdom of Crowds: Methods of Human Judgement Aggregation.(Submitted in Handbook of Human Computation. Springer, 2013).

[86] Mak, Brenda, Tung Bui, and Robert Blanning. Aggregating and updating experts' knowledge: an experimental evaluation of five classification techniques. Expert Systems with Applications 10.2 (1996): 233-241.

[87] McNeil, Alexander J., Rudiger Frey, and Paul Embrechts. Quantitative risk management: concepts, techniques, and tools. Princeton university press, 2010

[88] Mendel, Max B., and Thomas B. Sheridan. Filtering information from human experts. Systems, Man and Cybernetics, IEEE Transactions on 19.1 (1989): 6-16.

[89] Moon, Joo Hyun, and Chang Sun Kang. Use of fuzzy set theory in the aggregation of expert judgments. Annals of Nuclear Energy 26.6 (1999): 461-469.

[90] Morris, Peter A. Decision analysis expert use. Management Science 20.9 (1974): 1233-1241.

[91] Morris, Peter A. Combining expert judgments: A Bayesian approach. Management Science 23.7 (1977): 679-693.

[92] Morris, Peter A. An axiomatic approach to expert resolution. Management Science 29.1 (1983): 24-32.

[93] Muller, Peter, and Riten Mitra. Bayesian Nonparametric Inference Why and How. Bayesian Analysis 8.2 (2013): 269-302.

[94] Myung, In Jae, Sridhar Ramamoorti, and Andrew D. Bailey. Maximum entropy aggregation of expert predictions. Management Science 42.1 (1996): 1420-1436.

[95] Nelsen, Roger B. An introduction to copulas. Springer, 2006

[96] O'Hagan, Anthony, et al. Uncertain judgements: eliciting experts' probabilities. Wiley.com, 2006

[97] Okoli, Chitu, and Suzanne D. Pawlowski. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 42.1 (2004): 15-29.

[98] Otway, Harry, and Detlof Winterfeldt. Expert judgment in risk analysis and management: process, context, and pitfalls. Risk Analysis 12.1 (1992): 83-93.

[99] Pate-Cornell, M. Elisabeth. Uncertainties in risk analysis: Six levels of treatment. Reliability Engineering & System Safety 54.2 (1996): 95-111.

[100] Petrone, Sonia, Judith Rousseau, and Catia Scricciolo. Bayes and empirical Bayes: do they merge?. arXiv preprint arXiv:1204.1470 (2012).

[101] Picciotto, Sol. Is the International Tax System Fit for Purpose, Especially for Developing Countries?. (2013).

[102] Quigley, John and Walls, Lesley -2010 Reconciling experts opinion concerning the

value of testing using Bayesian networks : a bridge too far? In: 5th International ASRANet Conference, 14/06/2010 - 2010-06-16, Edinburgh.

[103] Quigley, John, and Lesley Walls. Mixing Bayes and empirical Bayes inference to anticipate the realization of engineering concerns about variant system designs. Reliability Engineering & System Safety 96.8 (2011): 933-941.

[104] Quigley, John and Wilson, Kevin and Bedford, Tim and Walls, Lesley -2013 Bayes linear Bayes graphical models in the design of optimal test strategies. International Journal of Performability Engineering. Vol 9, No. 6, pp 715-728.

[105] Quigley, John, et al. Merging expert and empirical data for rare event frequency estimation: Pool homogenisation for empirical Bayes models. Reliability Engineering & System Safety 96.6 (2011): 687-695.

[106] Robbins, Herbert. An empirical Bayes approach to statistics. Springer New York, 1985

[107] Rockafellar, R. Tyrrell. Lagrange multipliers and optimality. SIAM review (1993): 183-238.

[108] Rosenberg, Morris. Society and the adolescent self-image. Vol. 11. Princeton, NJ: Princeton university press, 1965.

[109] Sabou, M., Bontcheva, K., Scharl, A., & Fols, M. (2013). Games with a Purpose or Mechanised Labour? A Comparative Study.

[110] Sahin, Oz, and Sherif Mohamed. A spatial temporal decision framework for adaptation to sea level rise. Environmental Modelling & Software (2013).

[111] Salo, Ahti A. Interactive decision aiding for group decision support. European Journal of Operational Research 84.1 (1995): 134-149.

[112] Shirazi, Calvin Homayoon. Data-Informed Calibration and Aggregation of Expert Judgment in a Bayesian Framework. (2009).

[113] Surowiecki, James. The wisdom of crowds. Random House Digital, Inc., 2005

[114] Taylor, Angus Ellis, and William Robert Mann. Advanced calculus. New York: Wiley, 1983

[115] Tsyganok, Vitaliy V., Sergey V. Kadenko, and Oleg V. Andriichuk. Simulation of Expert Judgements for Testing the Methods of Information Processing in Decision-Making Support Systems. Journal of Automation and Information Sciences 43.12 (2011).

[116] Tukey, John W. Conclusions vs decisions. Technometrics 2.4 (1960): 423-433.

[117] Tukey, John W. The future of data analysis. The Annals of Mathematical Statistics 33.1 (1962): 1-67.

[118] v Mises, Richard. Probability, statistics, and truth. DoverPublications. com, 1957

[119] Vanicek, J., I. Vrana, and S. Aly. Fuzzy aggregation and averaging for group decision making: A generalization and survey. Knowledge-Based Systems 22.1 (2009): 79-84.

[120] Vose, David. Risk analysis: a quantitative guide. Wiley.com, 2008

[121] Vose, David. Quantitative risk analysis: a guide to Monte Carlo simulation modelling. Chichester: Wiley, 1996

[122] Walls, Lesley, and John Quigley. Building prior distributions to support Bayesian reliability growth modelling using expert judgement. Reliability Engineering & System Safety 74.2 (2001): 117-128.

[123] Winkler, Robert L. Combining probability distributions from dependent information sources. Management Science 27.4 (1981): 479-488.

[124] Wisse, Bram, Tim Bedford, and John Quigley. Expert judgement combination using moment methods. Reliability Engineering & System Safety 93.5 (2008): 675-686.

[125] Xu, Z. On consistency of the weighted geometric mean complex judgement matrix in AHP. European Journal of Operational Research 126.3 (2000): 683-687.

[126] You, Liangzhi, and Stanley Wood. An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems 90.1 (2006): 329-347.

[127] Zio, E., On the use of the analytic hierarchy process in the aggregation of expert judgements. Reliability Engineering & systems Safety 53.2(1996): 127-138.

[128] Montgomery, Douglas C., and Douglas C. Montgomery. Design and analysis of experiments. Vol. 7. New York: Wiley, 1984. 127-138.

[129] Zio, E., and G. E. Apostolakis. Accounting for expert-to-expert variability: a potential source of bias in performance assessments of high-level radioactive waste repositories. Annals of Nuclear Energy 24.1 (1997): 751-762.

Appendices

Appendix A

Dataset of an ageing bridge

A.1 Aggregated expert assessments for direct pull off test

The table below compares the aggregated assessments of the experts on their judgements on the direct pull-off test. The comparison is between the nonparametric methods, EB methods and simple average. The weights used for the experts for EB aggregation are : E1=0.2480957, E2=0.1980516, E3=0.2844636 and E4=0.2693891. The weights used for the experts for Non-parametric aggregation are: E1=0.05358375, E2=0.27847950, E3=0.37414749 and E4=0.29378926.

Table A.1: DPT - Test comparison results

Outcome	Condition	E1	E2	E3	E4	Average	EB	NP
FE	OLBFE	0.8	0.01	1	0.8	0.6525	0.178563205	0.163707673
JUD	OLBFE	0.8	0.5	0	0.7	0.5	0.154817005	0.096939808
NM	OLBFE	0.1	1	0	0.1	0.3	0.054225348	0.0783042

TF	OLBFE	0	0	0	0	0	0	0
FE	OLBPE	0.8	0.25	0	0.75	0.45	0.147720093	0.083207205
JUD	OLBPE	0.8	0.6	0.8	0.7	0.725	0.18854059	0.178731294
NM	OLBPE	0.1	1	0	0.1	0.3	0.054225348	0.0783042
TF	OLBPE	0	0	0	0	0	0	0
FE	OLBNE	0.7	0.35	0.65	0.7	0.6	0.160294698	0.1459562
JUD	OLBNE	0.7	0.25	0.0999	0.6	0.412475	0.131830843	0.080194848
NM	OLBNE	0.1	1	0	0.2	0.325	0.058355248	0.085648932
TF	OLBNE	0	0	0	0	0	0	0
FE	OLNBFE	0.8	0.01	0	0.85	0.415	0.143054605	0.073843167
JUD	OLNBFE	0.8	0.65	0	0.85	0.575	0.166508973	0.118399887
NM	OLNBFE	0.1	1	0	0.25	0.3375	0.060420198	0.089321298
TF	OLNBFE	0	0	0	0	0	0	0
FE	OLNBPE	0.8	0.25	0.05	0.85	0.4875	0.15372867	0.09522878
JUD	OLNBPE	0.8	0.45	0.8	0.7	0.6875	0.183043473	0.168288312
NM	OLNBPE	0.1	1	0	0.1	0.3	0.054225348	0.0783042
TF	OLNBPE	0	0	0	0	0	0	0
FE	OLNBNE	0.7	0.5	0.65	0.7	0.6375	0.165791815	0.156399181
JUD	OLNBNE	0.7	0	0.0999	0.7	0.374975	0.12679888	0.07013461
NM	OLNBNE	0.1	1	0	0.1	0.3	0.054225348	0.0783042
TF	OLNBNE	0	0	0	0	0	0	0

FE	RCALBFE	0.6	0.01	0	0.75	0.34	0.11202871	0.063819248
JUD	RCALBFE	0.6	0.8	0	0	0.35	0.110005945	0.063733463
NM	RCALBFE	0.1	1	0	0	0.275	0.050095448	0.070959469
TF	RCALBFE	0	0	0	0	0	0	0
FE	RCALBPE	0.6	0.2	0.1	0.75	0.4125	0.12274908	0.086400711
JUD	RCALBPE	0.6	0.15	0.65	0.5	0.475	0.13125741	0.116003168
NM	RCALBPE	0.1	1	0	0	0.275	0.050095448	0.070959469
TF	RCALBPE	0	0	0	0	0	0	0
FE	RCALBNE	0.6	0.35	0.75	0.75	0.6125	0.152669005	0.157642659
JUD	RCALBNE	0.6	0	0	0.7	0.325	0.109597285	0.059450683
NM	RCALBNE	0.1	1	0	0	0.275	0.050095448	0.070959469
TF	RCALBNE	0	0	0	0	0	0	0
FE	RCALFE	0.6	0.01	0	0.8	0.3525	0.11409366	0.067491613
JUD	RCALFE	0.6	0.5	0	0	0.275	0.09901171	0.0428475
NM	RCALFE	0.1	1	0	0	0.275	0.050095448	0.070959469
TF	RCALFE	0	0	0	0	0	0	0
FE	RCALPE	0.6	0.2	0.05	0.8	0.4125	0.122935353	0.085396233
JUD	RCALPE	0.6	0	0.7	0.6	0.475	0.13176887	0.117581762
NM	RCALPE	0.1	1	0	0	0.275	0.050095448	0.070959469
TF	RCALPE	0	0	0	0	0	0	0
FE	RCALNE	0.6	0.35	0.75	0.8	0.625	0.154733955	0.161315025

JUD	RCALNE	0.6	1	0	0.8	0.6	0.150374635	0.13641529
NM	RCALNE	0.1	0	0	0	0.025	0.013447998	0.001339594
TF	RCALNE	0	0	0	0	0	0	0
FE	RCULBFE	0.6	0.01	0	0.3	0.2275	0.09344416	0.030767956
JUD	RCULBFE	0.6	0.5	0.75	0.2	0.5125	0.135451673	0.127689617
NM	RCULBFE	0.1	0	0	0.1	0.05	0.017577898	0.008684325
TF	RCULBFE	0	1	0	0	0.25	0.03664745	0.069619875
FE	RCULPE	0.6	0.2	0	0.2	0.25	0.096277275	0.036651001
JUD	RCULPE	0.6	0.25	0.7	0.2	0.4375	0.124411133	0.105607805
NM	RCULPE	0.1	0	0	0.1	0.05	0.017577898	0.008684325
TF	RCULPE	0	1	0	0	0.25	0.03664745	0.069619875
FE	RCULNE	0.6	0.35	0	0.1	0.2625	0.097644493	0.03974925
JUD	RCULNE	0.6	0.05	0.5	0.1	0.3125	0.105437033	0.065631724
NM	RCULNE	0.1	0	0	0.05	0.0375	0.015512948	0.00501196
TF	RCULNE	0	1	0	0	0.25	0.03664745	0.069619875
FE	RCULNBFE	0.6	0	0	0.3	0.225	0.093077685	0.030071757
JUD	RCULNBFE	0.6	0.5	0.75	0.3	0.5375	0.139581573	0.135034349
NM	RCULNBFE	0.1	0	0	0.15	0.0625	0.019642848	0.012356691
TF	RCULNBFE	0	1	0	0	0.25	0.03664745	0.069619875
FE	RCULNBPE	0.6	0.2	0	0.3	0.275	0.100407175	0.043995732
JUD	RCULNBPE	0.6	0.25	0.65	0.3	0.45	0.126662355	0.108275693

NM	RCULNBPE	0.1	0	0	0.1	0.05	0.017577898	0.008684325
TF	RCULNBPE	0	1	0	0	0.25	0.03664745	0.069619875
FE	RCULNBNE	0.6	0.35	0	0.2	0.2875	0.101774393	0.047093982
JUD	RCULNBNE	0.6	0	0.65	0.2	0.3625	0.113370593	0.083525993
NM	RCULNBNE	0.1	0	0	0.1	0.05	0.017577898	0.008684325
TF	RCULNBNE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FBFE	0	0	0	0	0	0	0
JUD	FBFE	0	0	0.85	0	0.2125	0.031937518	0.079506342
NM	FBFE	0.2	0	0	0.4	0.15	0.043415595	0.032058114
TF	FBFE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FBPE	0	0	0	0	0	0	0
JUD	FBPE	0	0	0.45	0	0.1125	0.016908098	0.042091593
NM	FBPE	0.2	0	0	0.2	0.1	0.035155795	0.017368651
TF	FBPE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FBNE	0	0	0	0	0	0	0
JUD	FBNE	0	0	0	0	0	0	0
NM	FBNE	0.2	0	0	0	0.05	0.026895995	0.002679188
TF	FBNE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FNBFE	0	0	0	0	0	0	0
JUD	FNBFE	0	0	0.75	0	0.1875	0.028180163	0.070152654
NM	FNBFE	0.2	0	0	0.5	0.175	0.047545495	0.039402845

TF	FNBFE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FNBPE	0	0	0	0	0	0	0
JUD	FNBPE	0	0	0.75	0	0.1875	0.028180163	0.070152654
NM	FNBPE	0.2	0	0	0.2	0.1	0.035155795	0.017368651
TF	FNBPE	0	1	0	0	0.25	0.03664745	0.069619875
FE	FNBNE	0	0	0	0	0	0	0
JUD	FNBNE	0	0	0	0	0	0	0
NM	FNBNE	0.2	0	0	0	0.05	0.026895995	0.002679188
TF	FNBNE	0	1	0	0	0.25	0.03664745	0.069619875

A.2 Aggregated expert assessments for lab wire inspection test

The table below compares the aggregated assessments of the experts on their judgements on the labwire inspection test. The comparison is between the nonparametric methods, EB methods and simple average. The weights used for the experts for EB aggregation are : E1=0.2727272, E2=0.2300197, E3=0.2028609 and E4=0.2943922. The weights used for the experts for Non-parametric aggregation are: E1=0.09473991, E2=0.16327965, E3=0.30296004 and E4=0.43902039.

Table A.2: LWIT Test results

Outcome	Condition	E1	E2	E3	E4	Average	EB	NP

ALLOK	NNCNCNBW	0.95	1	0.95	0.9	0.95	0.9468	0.9362
SCWO	NNCNCNBW	0.35	0	1	0.4	0.4375	0.4161	0.5117
PO	NNCNCNBW	0.3	0	0	0.25	0.1375	0.1554	0.1382
SCPWO	NNCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
SCWCO	NNCNCNBW	0.2	0	0	0.3	0.125	0.1429	0.1507
SCWBWO	NNCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
PCO	NNCNCNBW	0.2	0	0	0.3	0.125	0.1429	0.1507
PBWO	NNCNCNBW	0.2	0	0	0.2	0.1	0.1134	0.1068
SCCBW	NNCNCNBW	0.2	0	0	0.2	0.1	0.1134	0.1068
PCBW	NNCNCNBW	0.2	0	0	0.2	0.1	0.1134	0.1068
SCPBW	NNCNCNBW	0.2	0	0	0.1	0.075	0.0840	0.0629
SCPC	NNCNCNBW	0.2	0	0	0.1	0.075	0.0840	0.0629
SCPCBW	NNCNCNBW	1	0	0	1	0.5	0.5671	0.5338
ALLOK	NNCNCBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	NNCNCBW	0.35	0	0.95	0.1	0.35	0.3176	0.3649
PO	NNCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCPWO	NNCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCWCO	NNCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCWBWO	NNCNCBW	0.5	0	1	0.3	0.45	0.4275	0.4820
PCO	NNCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	NNCNCBW	0.45	0	0	0.3	0.1875	0.2110	0.1743

SCCBW	NNCNCBW	0.45	0	0	0.3	0.1875	0.2110	0.1743
PCBW	NNCNCBW	0.45	0	0	0.2	0.1625	0.1816	0.1304
SCPBW	NNCNCBW	0.35	0	0	0.2	0.1375	0.1543	0.1210
SCPC	NNCNCBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCPCBW	NNCNCBW	1	0	0	1	0.5	0.5671	0.5338
ALLOK	NNCCPNBW	0	0	0.8	0	0.2	0.1623	0.2424
PO	NNCCPNBW	0.35	0	0.95	0.4	0.425	0.4059	0.4966
PO	NNCCPNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCPWO	NNCCPNBW	0.3	0	0	0.25	0.1375	0.1554	0.1382
SCWCO	NNCCPNBW	0.45	0	1	0.299	0.43725	0.4136	0.4769
SCWBWO	NNCCPNBW	0.35	0	0	0.3	0.1625	0.1838	0.1649
PCO	NNCCPNBW	0.45	0	0	0.3	0.1875	0.2110	0.1743
PBWO	NNCCPNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCCBW	NNCCPNBW	0.25	0	0	0.2	0.1125	0.1271	0.1115
PCBW	NNCCPNBW	0.25	0	0	0.2	0.1125	0.1271	0.1115
SCPBW	NNCCPNBW	0.25	0	0	0.2	0.1125	0.1271	0.1115
SCPC	NNCCPNBW	0.3	0	0	0.15	0.1125	0.1260	0.0943
SCPCBW	NNCCPNBW	1	0	0	1	0.5	0.5671	0.5338
ALLOK	NNCCPBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	NNCCPBW	0.35	0	0.95	0.2	0.375	0.3471	0.4088
PO	NNCCPBW	0.3	0	0	0.3	0.15	0.1701	0.1601

SCPWO	NNCCPBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCWCO	NNCCPBW	0.45	0	0	0.3	0.1875	0.2110	0.1743
SCWBWO	NNCCPBW	0.45	0	0	0.3	0.1875	0.2110	0.1743
PCO	NNCCPBW	0.4	0	0	0.4	0.2	0.2268	0.2135
PBWO	NNCCPBW	0.4	0	0	0.3	0.175	0.1974	0.1696
SCCBW	NNCCPBW	0.4	0	1	0.3	0.425	0.4003	0.4726
PCBW	NNCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCPBW	NNCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCPC	NNCCPBW	0.3	0	0	0.4	0.175	0.1996	0.2040
SCPCBW	NNCCPBW	1	0	0	1	0.5	0.5671	0.5338
ALLOK	NCNCNBW	0.9	0	0.85	0.85	0.65	0.6681	0.7159
SCWO	NCNCNBW	0.9	0.9	1	0.85	0.9125	0.9056	0.9083
PO	NCNCNBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCPWO	NCNCNBW	0.3	0.75	0	0.3	0.3375	0.3427	0.2826
SCWCO	NCNCNBW	0.3	0.2	0	0.25	0.1875	0.2014	0.1708
SCWBWO	NCNCNBW	0.25	0.2	0	0.25	0.175	0.1878	0.1661
PCO	NCNCNBW	0.2001	0	0	0.2	0.100025	0.1135	0.1068
PBWO	NCNCNBW	0.2001	0	0	0.2	0.100025	0.1135	0.1068
SCCBW	NCNCNBW	0.2001	0.2	0	0.2	0.150025	0.1595	0.1394
PCBW	NCNCNBW	0.2001	0	0	0.2	0.100025	0.1135	0.1068
SCPBW	NCNCNBW	0.25	0.2	0	0.2	0.1625	0.1731	0.1441

SCPC	NCNCNBW	0.25	1	0	0.2	0.3625	0.3571	0.2748
SCPCBW	NCNCNBW	1	0	0	1	0.5	0.5671	0.5338
ALLOK	NCNCBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	NCNCBW	0.6	0.8	0.95	0.8	0.7875	0.7759	0.8265
PO	NCNCBW	0.2001	0	0	0.35	0.137525	0.1576	0.1726
SCPWO	NCNCBW	0.35	0	0	0.35	0.175	0.1985	0.1868
SCWCO	NCNCBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCWBWO	NCNCBW	0.9	0.1	1	0.4	0.6	0.5891	0.5802
PCO	NCNCBW	0.2001	0	0	0.2	0.100025	0.1135	0.1068
PBWO	NCNCBW	0.4	0	0	0.35	0.1875	0.2121	0.1916
SCCBW	NCNCBW	0.4	0.1	0	0.4	0.225	0.2498	0.2298
PCBW	NCNCBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCPBW	NCNCBW	0.4	0.1	0	0.4	0.225	0.2498	0.2298
SCPC	NCNCBW	0.35	0.1	0	0.4	0.2125	0.2362	0.2251
PO	NCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	NCCPNBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	NCCPNBW	0.3	0.8	0.95	0.2	0.5625	0.5174	0.5347
PO	NCCPNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCPWO	NCCPNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCWCO	NCCPNBW	0.9	0.5	1	0.8	0.8	0.7988	0.8211
SCWBWO	NCCPNBW	0.4	0	0	0.4	0.2	0.2268	0.2135

PCO	NCCPNBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
PBWO	NCCPNBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCCBW	NCCPNBW	0.4	0.5	0	0.3	0.3	0.3124	0.2512
PCBW	NCCPNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCPBW	NCCPNBW	0.3	0.15	0	0.2	0.1625	0.1752	0.1407
SCPC	NCCPNBW	0.4	0.15	0	0.3	0.2125	0.2319	0.1941
SCPCBW	NCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	NCCPBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	NCCPBW	0.3	0.6	0.9	0.4	0.55	0.5202	0.5747
PO	NCCPBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCPWO	NCCPBW	0.35	0.5	0	0.4	0.3125	0.3282	0.2904
SCWCO	NCCPBW	0.4	0.25	0.5	0.4	0.3875	0.3858	0.4058
SCWBWO	NCCPBW	0.4	0.05	0.5	0.3	0.3125	0.3103	0.3292
PCO	NCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
PBWO	NCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCCBW	NCCPBW	0.9	0	1	0.5	0.6	0.5955	0.6077
PCBW	NCCPBW	0.4	0.9	0	0.6	0.475	0.4927	0.4483
SCPBW	NCCPBW	0.4	0	0	0.3	0.175	0.1974	0.1696
SCPC	NCCPBW	0.4	0.1	0	0.3	0.2	0.2204	0.1859
SCPCBW	NCCPBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWONCNCNBW	0	0	0.8	0	0.2	0.1623	0.2424

SCWO	IWONCNCNBW	0.3	0	0.5	0.4	0.3	0.3010	0.3555
PO	IWONCNCNBW	0.9	0.25	0.9	0.701	0.68775	0.6919	0.7065
SCPWO	IWONCNCNBW	0.3	0.2	1	0.4	0.475	0.4484	0.5396
SCWCO	IWONCNCNBW	0.25	0.5	0	0.3	0.2625	0.2715	0.2370
SCWBWO	IWONCNCNBW	0.2001	0	0	0.3	0.125025	0.1429	0.1507
PCO	IWONCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
PBWO	IWONCNCNBW	0.25	0	0	0.2	0.1125	0.1271	0.1115
SCCBW	IWONCNCNBW	0.25	0.5	0	0.2	0.2375	0.2421	0.1931
PCBW	IWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPBW	IWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPC	IWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPCBW	IWONCNCNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWONCNCBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	IWONCNCBW	0.3	0	0.55	0.3	0.2875	0.2817	0.3268
PO	IWONCNCBW	0.4	0	0.5	0.4	0.325	0.3283	0.3650
SCPWO	IWONCNCBW	0.4	0	0.5	0.3	0.3	0.2988	0.3211
SCWCO	IWONCNCBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCWBWO	IWONCNCBW	0.4	0	0.0501	0.3	0.187525	0.2076	0.1848
PCO	IWONCNCBW	0.4	0	0	0.3	0.175	0.1974	0.1696
PBWO	IWONCNCBW	0.9	0	0.75	0.95	0.65	0.6773	0.7296
SCCBW	IWONCNCBW	0.35	0.5	0	0.4	0.3125	0.3282	0.2904

PCBW	IWONCNCBW	0.4	0.6	0	0.499	0.37475	0.3940	0.3549
SCPBW	IWONCNCBW	0.4	0.05	1	0.4	0.4625	0.4412	0.5246
SCPC	IWONCNCBW	0.35	0.05	0	0.3	0.175	0.1953	0.1730
SCPCBW	IWONCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWONCCPNBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	IWONCCPNBW	0.3	0	0.55	0.2	0.2625	0.2523	0.2829
PO	IWONCCPNBW	0.4	0	0.5	0.1	0.25	0.2400	0.2333
SCPWO	IWONCCPNBW	0.4	0	0.5	0.2	0.275	0.2694	0.2772
SCWCO	IWONCCPNBW	0.4	0	0	0.6	0.25	0.2857	0.3013
SCWBWO	IWONCCPNBW	0.3	0	0.85	0.3	0.3625	0.3426	0.4176
PCO	IWONCCPNBW	0.9	0.1	0	0.7	0.425	0.4745	0.4089
PBWO	IWONCCPNBW	0.4	0	0	0.3	0.175	0.1974	0.1696
SCCBW	IWONCCPNBW	0.3	0.8	0	0.4	0.375	0.3836	0.3347
PCBW	IWONCCPNBW	0.35	0.8	0	0.4	0.3875	0.3972	0.3394
SCPBW	IWONCCPNBW	0.3	0.15	0	0.3	0.1875	0.2046	0.1846
SCPC	IWONCCPNBW	0.4	0.15	1	0.2	0.4375	0.4053	0.4532
SCPCBW	IWONCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWONCCPBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	IWONCCPBW	0.3	0	0.55	0.2	0.2625	0.2523	0.2829
PO	IWONCCPBW	0.4	0.1	0.5	0.3	0.325	0.3218	0.3374
SCPWO	IWONCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088

SCWCO	IWONCCPBW	0.35	0	0	0.3	0.1625	0.1838	0.1649
SCWBWO	IWONCCPBW	0.4	0	0	0.4	0.2	0.2268	0.2135
PCO	IWONCCPBW	0.45	0	0.5	0.3	0.3125	0.3125	0.3258
PBWO	IWONCCPBW	0.45	0	1	0.3	0.4375	0.4139	0.4773
SCCBW	IWONCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
PCBW	IWONCCPBW	0.9	0.5	0	0.85	0.5625	0.6107	0.5401
SCPBW	IWONCCPBW	0.35	0.2	0	0.4	0.2375	0.2592	0.2414
SCPC	IWONCCPBW	0.35	0.2	0	0.2	0.1875	0.2003	0.1536
SCPCBW	IWONCCPBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWOCNCNBW	0	0	0.7	0	0.175	0.1420	0.2121
PO	IWOCNCNBW	0.3	0.8	0.95	0.1	0.5375	0.4880	0.4908
PO	IWOCNCNBW	0.3	0	0.8	0.1	0.3	0.2735	0.3147
SCPWO	IWOCNCNBW	0.9	0.3	1	0.95	0.7875	0.7970	0.8543
SCWCO	IWOCNCNBW	0.3	0.1	0	0.2	0.15	0.1637	0.1326
SCWBWO	IWOCNCNBW	0.3	0.1	0	0.2	0.15	0.1637	0.1326
PCO	IWOCNCNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	IWOCNCNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCCBW	IWOCNCNBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
PCBW	IWOCNCNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCPBW	IWOCNCNBW	0.4	0.1	0	0.2	0.175	0.1910	0.1420
SCPC	IWOCNCNBW	0.4	0.3	0	0.2	0.225	0.2370	0.1747

SCPCBW	IWOCNCNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWOCNCBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	IWOCNCBW	0.3	0.8	0.85	0.3	0.5625	0.5266	0.5483
PO	IWOCNCBW	0.3	0	0.85	0.3	0.3625	0.3426	0.4176
PO	IWOCNCBW	0.4	0.3	0.4	0.4	0.375	0.3770	0.3837
SCWCO	IWOCNCBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
PO	IWOCNCBW	0.4	0.1	0.0499	0.4	0.237475	0.2600	0.2449
PCO	IWOCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	IWOCNCBW	0.4	0.05	0.5	0.4	0.3375	0.3398	0.3731
SCCBW	IWOCNCBW	0.35	0.1	0	0.35	0.2	0.2215	0.2031
PCBW	IWOCNCBW	0.35	0	0	0.35	0.175	0.1985	0.1868
SCPBW	IWOCNCBW	0.9	0.1	1	0.9	0.725	0.7363	0.7997
SCPC	IWOCNCBW	0.35	0.3	0	0.35	0.25	0.2675	0.2358
SCPCBW	IWOCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWOCCPNBW	0	0	0.6501	0	0.162525	0.1319	0.1970
SCWO	IWOCCPNBW	0.3	0.8	0.85	0.3	0.5625	0.5266	0.5483
PO	IWOCCPNBW	0.3	0	0.75	0.3	0.3375	0.3223	0.3873
SCPWO	IWOCCPNBW	0.4	0.3	0.6	0.4	0.425	0.4176	0.4443
SCWCO	IWOCCPNBW	0.4	0.1	0.05	0.4	0.2375	0.2600	0.2450
SCWBWO	IWOCCPNBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
PCO	IWOCCPNBW	0.4	0.05	0.5	0.4	0.3375	0.3398	0.3731

PBWO	IWOCCPNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCCBW	IWOCCPNBW	0.4	0.1	0	0.4	0.225	0.2498	0.2298
PCBW	IWOCCPNBW	0.4	0	0	0.4	0.2	0.2268	0.2135
SCPBW	IWOCCPNBW	0.4	0.1	0	0.4	0.225	0.2498	0.2298
SCPC	IWOCCPNBW	0.9	0.3	1	0.9	0.775	0.7823	0.8323
SCPCBW	IWOCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IWOCCPBW	0	0	0.6501	0	0.162525	0.1319	0.1970
SCWO	IWOCCPBW	0.25	0.8	0.8	0.2	0.5125	0.4734	0.4845
PO	IWOCCPBW	0.25	0	0.6	0.2	0.2625	0.2488	0.2933
SCPWO	IWOCCPBW	0.3	0.3	0.4	0.4	0.35	0.3497	0.3742
SCWCO	IWOCCPBW	0.3	0.1	0.05	0.4	0.2125	0.2327	0.2355
SCWBWO	IWOCCPBW	0.3	0.1	0.0499	0.3	0.187475	0.2033	0.1916
PO	IWOCCPBW	0.3	0.05	0.5	0.4	0.3125	0.3125	0.3637
PBWO	IWOCCPBW	0.3	0.05	0.5	0.3	0.2875	0.2831	0.3198
SCCBW	IWOCCPBW	0.35	0.1	0.0499	0.3	0.199975	0.2169	0.1963
PCBW	IWOCCPBW	0.35	0	1	0.4	0.4375	0.4161	0.5117
SCPBW	IWOCCPBW	0.35	0.1	0	0.4	0.2125	0.2362	0.2251
SCPC	IWOCCPBW	0.35	0.3	0	0.4	0.2625	0.2822	0.2578
SCPCBW	IWOCCPBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWONCNCNBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	EWONCNCNBW	0.35	0	0.95	0.4	0.425	0.4059	0.4966

PO	EWONCNCNBW	0.9	0.25	1	0.4	0.6375	0.6236	0.6047
SCPWO	EWONCNCNBW	0.35	0.2	0	0.399	0.23725	0.2589	0.2410
SCWCO	EWONCNCNBW	0.25	0.5	0	0.3	0.2625	0.2715	0.2370
SCWBWO	EWONCNCNBW	0.2	0	0	0.3	0.125	0.1429	0.1507
PCO	EWONCNCNBW	0.25	0	0	0.299	0.13725	0.1562	0.1550
PBWO	EWONCNCNBW	0.25	0	0	0.2	0.1125	0.1271	0.1115
SCCBW	EWONCNCNBW	0.25	0.5	0	0.2	0.2375	0.2421	0.1931
PCBW	EWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPBW	EWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPC	EWONCNCNBW	0.25	0.05	0	0.2	0.125	0.1386	0.1197
SCPCBW	EWONCNCNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWONCNCBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	EWONCNCBW	0.35	0	0.9	0.3	0.3875	0.3663	0.4375
PO	EWONCNCBW	0.35	0	0.05	0.4	0.2	0.2234	0.2239
SCPWO	EWONCNCBW	0.35	0	0	0.3	0.1625	0.1838	0.1649
SCWCO	EWONCNCBW	0.3	0	0	0.4	0.175	0.1996	0.2040
SCWBWO	EWONCNCBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
PCO	EWONCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	EWONCNCBW	0.9	0	0.05	0.95	0.475	0.5353	0.5175
SCCBW	EWONCNCBW	0.3	0.5	0	0.4	0.3	0.3146	0.2857
PCBW	EWONCNCBW	0.35	0.6	0	0.499	0.36225	0.3804	0.3502

SCPBW	EWONCNCBW	0.35	0.05	1	0.4	0.45	0.4276	0.5199
SCPC	EWONCNCBW	0.35	0.05	0	0.3	0.175	0.1953	0.1730
SCPCBW	EWONCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWONCCPNBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	EWONCCPNBW	0.35	0	0.9	0.2	0.3625	0.3369	0.3936
PO	EWONCCPNBW	0.35	0	0.05	0.1	0.125	0.1350	0.0922
SCPWO	EWONCCPNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCWCO	EWONCCPNBW	0.3	0	0	0.6	0.225	0.2585	0.2918
SCWBWO	EWONCCPNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
PCO	EWONCCPNBW	0.9	0	0.75	0.7	0.5875	0.6037	0.6198
PBWO	EWONCCPNBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
SCCBW	EWONCCPNBW	0.25	0	0	0.4	0.1625	0.1859	0.1993
PCBW	EWONCCPNBW	0.3	0.8	0	0.4	0.375	0.3836	0.3347
SCPBW	EWONCCPNBW	0.25	0.8	0	0.3	0.3375	0.3405	0.2860
SCPC	EWONCCPNBW	0.35	0.15	1	0.2	0.425	0.3917	0.4484
SCPCBW	EWONCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWONCCPBW	0	0	0.8	0	0.2	0.1623	0.2424
SCWO	EWONCCPBW	0.3	0	0.8	0.2	0.325	0.3030	0.3586
PO	EWONCCPBW	0.35	0.1	0.05	0.3	0.2	0.2169	0.1963
SCPWO	EWONCCPBW	0.3	0	0	0.4	0.175	0.1996	0.2040
SCWCO	EWONCCPBW	0.3	0	0	0.3	0.15	0.1701	0.1601

SCWBWO	EWONCCPBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
PCO	EWONCCPBW	0.4	0	0.4	0.3	0.275	0.2786	0.2908
PBWO	EWONCCPBW	0.4	0	0.4	0.3	0.275	0.2786	0.2908
SCCBW	EWONCCPBW	0.3	0	0	0.4	0.175	0.1996	0.2040
PCBW	EWONCCPBW	0.9	0.5	1	0.85	0.8125	0.8136	0.8430
SCPBW	EWONCCPBW	0.3	0.2	0	0.4	0.225	0.2456	0.2367
SCPC	EWONCCPBW	0.3	0.2	0	0.2	0.175	0.1867	0.1489
SCPCBW	EWONCCPBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWOCNCNBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	EWOCNCNBW	0.3	0.8	0.95	0.1	0.5375	0.4880	0.4908
PO	EWOCNCNBW	0.3	0	0.95	0.1	0.3375	0.3040	0.3601
PO	EWOCNCNBW	0.9	0.3	1	0.95	0.7875	0.7970	0.8543
PO	EWOCNCNBW	0.25	0.1	0	0.2	0.1375	0.1501	0.1278
SCWBWO	EWOCNCNBW	0.25	0.1	0	0.2	0.1375	0.1501	0.1278
PO	EWOCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
PBWO	EWOCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
SCCBW	EWOCNCNBW	0.25	0.1	0	0.3	0.1625	0.1795	0.1717
PCBW	EWOCNCNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554
SCPBW	EWOCNCNBW	0.3	0.1	0	0.2	0.15	0.1637	0.1326
SCPC	EWOCNCNBW	0.3	0.3	0	0.2	0.2	0.2097	0.1652
SCPCBW	EWOCNCNBW	1	1	0	1	0.75	0.7971	0.6970

ALLOK	EWOCNCBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	EWOCNCBW	0.25	0.8	0.6	0.3	0.4875	0.4622	0.4678
PO	EWOCNCBW	0.25	0	0.2	0.3	0.1875	0.1971	0.2160
SCPWO	EWOCNCBW	0.35	0.3	0.4	0.4	0.3625	0.3634	0.3789
SCWCO	EWOCNCBW	0.25	0.1	0.0499	0.3	0.174975	0.1896	0.1868
SCWBWO	EWOCNCBW	0.35	0.1	0	0.4	0.2125	0.2362	0.2251
PCO	EWOCNCBW	0.25	0	0.1	0.3	0.1625	0.1768	0.1857
PBWO	EWOCNCBW	0.35	0.05	0	0.4	0.2	0.2247	0.2169
SCCBW	EWOCNCBW	0.3	0.1	0	0.35	0.1875	0.2079	0.1984
PCBW	EWOCNCBW	0.3	0	0	0.35	0.1625	0.1849	0.1821
SCPBW	EWOCNCBW	0.9	0.1	0	0.9	0.475	0.5334	0.4967
SCPC	EWOCNCBW	0.3	0.3	1	0.35	0.4875	0.4567	0.5340
SCPCBW	EWOCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWOCCPNBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	EWOCCPNBW	0.25	0.8	0.6	0.3	0.4875	0.4622	0.4678
PO	EWOCCPNBW	0.25	0	0.2	0.3	0.1875	0.1971	0.2160
SCPWO	EWOCCPNBW	0.35	0.3	0.4	0.4	0.3625	0.3634	0.3789
SCWCO	EWOCCPNBW	0.35	0.1	0.0499	0.4	0.224975	0.2463	0.2402
SCWBWO	EWOCCPNBW	0.25	0.1	0	0.3	0.1625	0.1795	0.1717
PCO	EWOCCPNBW	0.25	0.05	0.1	0.4	0.2	0.2177	0.2378
PBWO	EWOCCPNBW	0.25	0	0	0.3	0.1375	0.1565	0.1554

SCCBW	EWOCCPNBW	0.35	0.1	0	0.4	0.2125	0.2362	0.2251
PCBW	EWOCCPNBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
SCPBW	EWOCCPNBW	0.35	0.1	0	0.4	0.2125	0.2362	0.2251
SCPC	EWOCCPNBW	0.9	0.3	1	0.9	0.775	0.7823	0.8323
SCPCBW	EWOCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	EWOCCPBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	EWOCCPBW	0.2	0.8	0.55	0.2	0.4375	0.4090	0.4040
PO	EWOCCPBW	0.2	0	0.2	0.2	0.15	0.1540	0.1673
SCPWO	EWOCCPBW	0.25	0.3	0.3499	0.4	0.324975	0.3259	0.3543
SCWCO	EWOCCPBW	0.25	0.1	0.0499	0.4	0.199975	0.2191	0.2307
SCWBWO	EWOCCPBW	0.25	0.1	0.0499	0.3	0.174975	0.1896	0.1868
PCO	EWOCCPBW	0.25	0.05	0.1501	0.4	0.212525	0.2279	0.2529
PBWO	EWOCCPBW	0.25	0.05	0.1501	0.3	0.187525	0.1984	0.2090
SCCBW	EWOCCPBW	0.3	0.1	0.0499	0.3	0.187475	0.2033	0.1916
PCBW	EWOCCPBW	0.3	0	0.55	0.4	0.3125	0.3111	0.3707
SCPBW	EWOCCPBW	0.3	0.1	0.15	0.4	0.2375	0.2530	0.2658
SCPC	EWOCCPBW	0.3	0.3	0.5	0.4	0.375	0.3700	0.4045
SCPCBW	EWOCCPBW	1	1	1	1	1	1.0000	1.0000
ALLOK	IEWNCNCNBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	IEWNCNCNBW	0.35	0	0.85	0.3	0.375	0.3562	0.4224
PO	IEWNCNCNBW	0.9	0.25	0.95	0.8	0.725	0.7312	0.7651

SCPWO	IEWNCNCNBW	0.4	0.2	1	0.4	0.5	0.4757	0.5491
SCWCO	IEWNCNCNBW	0.3	0.5	0	0.2	0.25	0.2557	0.1979
SCWBWO	IEWNCNCNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
PCO	IEWNCNCNBW	0.3	0	0	0.4	0.175	0.1996	0.2040
PBWO	IEWNCNCNBW	0.3	0	0	0.1	0.1	0.1113	0.0723
SCCBW	IEWNCNCNBW	0.3	0.5	0	0.2	0.25	0.2557	0.1979
PCBW	IEWNCNCNBW	0.3	0.05	0	0.2	0.1375	0.1522	0.1244
SCPBW	IEWNCNCNBW	0.3	0.05	0	0.2	0.1375	0.1522	0.1244
SCPC	IEWNCNCNBW	0.3	0.05	0	0.2	0.1375	0.1522	0.1244
PO	IEWNCNCNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWNCNCBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	IEWNCNCBW	0.35	0	0.7	0.3	0.3375	0.3258	0.3769
PO	IEWNCNCBW	0.4	0	0.4	0.4	0.3	0.3080	0.3347
SCPWO	IEWNCNCBW	0.4	0	0.5	0.4	0.325	0.3283	0.3650
SCWCO	IEWNCNCBW	0.35	0	0	0.3	0.1625	0.1838	0.1649
SCWBWO	IEWNCNCBW	0.3	0	0.1	0.3	0.175	0.1904	0.1904
PCO	IEWNCNCBW	0.35	0	0	0.4	0.1875	0.2132	0.2088
PBWO	IEWNCNCBW	0.9	0	0	0.8	0.425	0.4810	0.4365
SCCBW	IEWNCNCBW	0.35	0.5	0	0.4	0.3125	0.3282	0.2904
PCBW	IEWNCNCBW	0.4	0.6	0	0.3	0.325	0.3354	0.2676
SCPBW	IEWNCNCBW	0.4	0.05	1	0.3	0.4375	0.4118	0.4807

SCPC	IEWNCNCBW	0.4	0.05	0	0.3	0.1875	0.2089	0.1778
SCPCBW	IEWNCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWNCCPNBW	0	0	0.75	0	0.1875	0.1521	0.2272
SCWO	IEWNCCPNBW	0.35	0	0.7	0.2	0.3125	0.2963	0.3330
PO	IEWNCCPNBW	0.4	0	0.4	0.3	0.275	0.2786	0.2908
SCPWO	IEWNCCPNBW	0.35	0	0.5	0.4	0.3125	0.3146	0.3602
SCWCO	IEWNCCPNBW	0.35	0	0.0499	0.4	0.199975	0.2233	0.2239
SCWBWO	IEWNCCPNBW	0.3	0	0	0.4	0.175	0.1996	0.2040
PCO	IEWNCCPNBW	0.9	0.1	0.101	0.8	0.47525	0.5245	0.4834
PBWO	IEWNCCPNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
SCCBW	IEWNCCPNBW	0.3	0.8	0	0.3	0.35	0.3542	0.2908
PCBW	IEWNCCPNBW	0.35	0.8	0	0.4	0.3875	0.3972	0.3394
SCPBW	IEWNCCPNBW	0.3	0.15	0	0.4	0.2125	0.2341	0.2285
SCPC	IEWNCCPNBW	0.35	0.15	1	0.4	0.475	0.4506	0.5362
SCPCBW	IEWNCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWNCCPBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	IEWNCCPBW	0.3	0	0.65	0.3	0.3125	0.3020	0.3571
PO	IEWNCCPBW	0.35	0.1	0.25	0.3	0.25	0.2575	0.2569
SCPWO	IEWNCCPBW	0.35	0	0.35	0.3	0.25	0.2548	0.2709
SCWCO	IEWNCCPBW	0.35	0	0.05	0.4	0.2	0.2234	0.2239
SCWBWO	IEWNCCPBW	0.35	0	0.05	0.5	0.225	0.2528	0.2678

PCO	IEWNCCPBW	0.4	0	0.15	0.3	0.2125	0.2278	0.2150
PBWO	IEWNCCPBW	0.4	0	0.15	0.3	0.2125	0.2278	0.2150
SCCBW	IEWNCCPBW	0.35	0	0.05	0.3	0.175	0.1939	0.1800
PCBW	IEWNCCPBW	0.9	0.5	0.65	0.85	0.725	0.7426	0.7370
SCPBW	IEWNCCPBW	0.35	0.2	0.15	0.5	0.3	0.3191	0.3308
SCPC	IEWNCCPBW	0.35	0.2	0.45	0.4	0.35	0.3505	0.3778
SCPCBW	IEWNCCPBW	1	1	1	1	1	1.0000	1.0000
ALLOK	IEWCNCNBW	0	0	0.7	0	0.175	0.1420	0.2121
SCWO	IEWCNCNBW	0.3	0.8	0.9	0.3	0.575	0.5367	0.5634
PO	IEWCNCNBW	0.3	0	0.9	0.3	0.375	0.3527	0.4328
SCPWO	IEWCNCNBW	0.9	0.3	1	0.8	0.75	0.7528	0.7884
SCWCO	IEWCNCNBW	0.3	0.1	0	0.2	0.15	0.1637	0.1326
SCWBWO	IEWCNCNBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
PCO	IEWCNCNBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	IEWCNCNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCCBW	IEWCNCNBW	0.3	0.1	0	0.2	0.15	0.1637	0.1326
PCBW	IEWCNCNBW	0.3	0	0	0.5	0.2	0.2290	0.2479
SCPBW	IEWCNCNBW	0.35	0.1	0	0.2	0.1625	0.1773	0.1373
SCPC	IEWCNCNBW	0.35	0.3	0	0.2	0.2125	0.2233	0.1699
SCPCBW	IEWCNCNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWCNCBW	0	0	0.65	0	0.1625	0.1319	0.1969

SCWO	IEWCNCBW	0.3	0.8	0.6	0.3	0.5	0.4759	0.4725
PO	IEWCNCBW	0.3	0	0.3	0.2	0.2	0.2016	0.2071
SCPWO	IEWCNCBW	0.4	0.3	0.3	0.4	0.35	0.3567	0.3534
SCWCO	IEWCNCBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
SCWBWO	IEWCNCBW	0.4	0.1	0.15	0.4	0.2625	0.2803	0.2753
PCO	IEWCNCBW	0.3	0	0	0.3	0.15	0.1701	0.1601
PBWO	IEWCNCBW	0.4	0.05	0	0.3	0.1875	0.2089	0.1778
SCCBW	IEWCNCBW	0.35	0.1	0	0.2	0.1625	0.1773	0.1373
PCBW	IEWCNCBW	0.35	0	0	0.2	0.1375	0.1543	0.1210
SCPBW	IEWCNCBW	0.9	0.1	1	0.8	0.7	0.7068	0.7558
SCPC	IEWCNCBW	0.35	0.3	0	0.2	0.2125	0.2233	0.1699
SCPCBW	IEWCNCBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWCCPNBW	0	0	0.65	0	0.1625	0.1319	0.1969
PO	IEWCCPNBW	0.3	0.8	0.6	0.2	0.475	0.4464	0.4286
PO	IEWCCPNBW	0.3	0	0.3	0.2	0.2	0.2016	0.2071
SCPWO	IEWCCPNBW	0.4	0.3	0.3	0.4	0.35	0.3567	0.3534
SCWCO	IEWCCPNBW	0.4	0.1	0.05	0.3	0.2125	0.2306	0.2011
SCWBWO	IEWCCPNBW	0.3	0.1	0	0.3	0.175	0.1931	0.1765
PCO	IEWCCPNBW	0.4	0.05	0.15	0.3	0.225	0.2393	0.2232
PBWO	IEWCCPNBW	0.3	0	0	0.2	0.125	0.1407	0.1162
SCCBW	IEWCCPNBW	0.4	0.1	0	0.3	0.2	0.2204	0.1859

PCBW	IEWCCPNBW	0.4	0	0	0.3	0.175	0.1974	0.1696
SCPBW	IEWCCPNBW	0.4	0.1	0	0.4	0.225	0.2498	0.2298
PO	IEWCCPNBW	0.9	0.3	1	0.8	0.75	0.7528	0.7884
SCPCBW	IEWCCPNBW	1	1	0	1	0.75	0.7971	0.6970
ALLOK	IEWCCPBW	0	0	0.55	0	0.1375	0.1116	0.1666
SCWO	IEWCCPBW	0.25	0.8	0.5	0.2	0.4375	0.4125	0.3936
PO	IEWCCPBW	0.25	0	0.25	0.2	0.175	0.1778	0.1872
SCPWO	IEWCCPBW	0.3	0.3	0.4	0.2	0.3	0.2908	0.2864
SCWCO	IEWCCPBW	0.3	0.1	0.0499	0.2	0.162475	0.1738	0.1477
SCWBWO	IEWCCPBW	0.3	0.1	0.0499	0.3	0.187475	0.2033	0.1916
PCO	IEWCCPBW	0.3	0.05	0.15	0.3	0.2	0.2121	0.2137
PBWO	IEWCCPBW	0.3	0.05	0.15	0.4	0.225	0.2415	0.2576
SCCBW	IEWCCPBW	0.35	0.1	0.0499	0.2	0.174975	0.1875	0.1524
PCBW	IEWCCPBW	0.35	0	0.6	0.3	0.3125	0.3055	0.3466
SCPBW	IEWCCPBW	0.35	0.1	0.1	0.3	0.2125	0.2271	0.2115
SCPC	IEWCCPBW	0.35	0.3	0.6	0.3	0.3875	0.3745	0.3956
SCPCBW	IEWCCPBW	1	1	1	1	1	1.0000	1.0000

A.3 Aggregated expert assessments for on-site inspection test

The table below compares the aggregated assessments of the experts on their judgements on onsite inspection test. The comparison is between the nonparametric methods, EB methods and simple average. The weights used for the experts for EB aggregation are : E1=0.2571549, E2=0.2295041, E3=0.2325712 and E4=0.2807698. The weights used for the experts for Non-parametric aggregation are: E1=0.006165874, E2=0.273986858, E3=0.344075389 and E4=0.375771879.

Table A.3: ONS Test comparison results

Outcome	Condition	E1	E2	E3	E4	Average	EB	NP
ALLOK	NNCNCNBW	0.98	1	1	0.9	0.97	0.2581	0.2406
SCWO	NNCNCNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PO	NNCNCNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCPWO	NNCNCNBW	0.2	0	0	0.1	0.08	0.0199	0.0097
SCWCO	NNCNCNBW	0.15	0	0	0.1	0.06	0.0167	0.0096
SCWBWO	NNCNCNBW	0.15	0	0	0.2	0.09	0.0237	0.0190
PCO	NNCNCNBW	0.15	0	0	0.2	0.09	0.0237	0.0190
PBWO	NNCNCNBW	0.15	0	1	0.2	0.34	0.0818	0.1050
SCCBW	NNCNCNBW	0.1	0	0	0.1	0.05	0.0134	0.0095
PCBW	NNCNCNBW	0.1	0	0	0.1	0.05	0.0134	0.0095
SCPBW	NNCNCNBW	0.1	0	0	0.1	0.05	0.0134	0.0095
SCPC	NNCNCNBW	0.1	1	0	0.1	0.30	0.0872	0.0780
SCPCBW	NNCNCNBW	0	0	0	1	0.25	0.0702	0.0939
ALLOK	NNCNCBW	1	1	1	1	1.00	0.2664	0.2500
SCWO	NNCNCBW	0.3	0	0	0.3	0.15	0.0403	0.0286

PO	NNCNCBW	0.25	1	1	0.2	0.61	0.1620	0.1737
SCPWO	NNCNCBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCWCO	NNCNCBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCWBWO	NNCNCBW	0.45	1	0	0.3	0.44	0.1237	0.0974
PCO	NNCNCBW	0.2	1	0	0.2	0.35	0.1007	0.0876
PBWO	NNCNCBW	0.4	0	0	0.3	0.18	0.0468	0.0288
SCCBW	NNCNCBW	0.35	0	0	0.2	0.14	0.0365	0.0193
PCBW	NNCNCBW	0.35	0	1	0.2	0.39	0.0947	0.1053
SCPBW	NNCNCBW	0.35	0	0	0.2	0.14	0.0365	0.0193
SCPC	NNCNCBW	0.3	0	0	0.2	0.13	0.0333	0.0193
SCPCBW	NNCNCBW	1	0	0	1	0.50	0.1345	0.0955
ALLOK	NNCCPNBW	0	0	1	0	0.25	0.0581	0.0860
SCWO	NNCCPNBW	0.3	0	0	0.2	0.13	0.0333	0.0193
PO	NNCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCPWO	NNCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCWCO	NNCCPNBW	0.4	0	0	0.4	0.20	0.0538	0.0382
SCWBWO	NNCCPNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PCO	NNCCPNBW	0.4	0	0	0.4	0.20	0.0538	0.0382
PBWO	NNCCPNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCCBW	NNCCPNBW	0.2	0	0	0.201	0.10	0.0270	0.0192
PCBW	NNCCPNBW	0.2	0	0	0.201	0.10	0.0270	0.0192

SCPBW	NNCCPNBW	0.2	0	0	0.101	0.08	0.0199	0.0098
SCPC	NNCCPNBW	0.2	0	0	0.101	0.08	0.0199	0.0098
SCPCBW	NNCCPNBW	1	0	0	1	0.50	0.1345	0.0955
ALLOK	NNCCPBW	0	0	1	0	0.25	0.0581	0.0860
SCWO	NNCCPBW	0.3	0	0	0.201	0.13	0.0334	0.0193
PO	NNCCPBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCPWO	NNCCPBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCWCO	NNCCPBW	0.4	0	0	0.4	0.20	0.0538	0.0382
SCWBWO	NNCCPBW	0.4	0	0	0.4	0.20	0.0538	0.0382
PCO	NNCCPBW	0.35	0.001	0	0.3	0.16	0.0436	0.0288
PBWO	NNCCPBW	0.35	1	0	0.4	0.44	0.1243	0.1066
SCCBW	NNCCPBW	0.35	0	0	0.3	0.16	0.0436	0.0287
PCBW	NNCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCPBW	NNCCPBW	0.3	0	0	0.201	0.13	0.0334	0.0193
SCPC	NNCCPBW	0.25	0	0	0.201	0.11	0.0302	0.0193
SCPCBW	NNCCPBW	1	0	0	1	0.50	0.1345	0.0955
ALLOK	NCNCNBW	0.9	0	0	0.95	0.46	0.1245	0.0906
SCWO	NCNCNBW	0.9	0.9	1	0.801	0.90	0.2386	0.2243
PO	NCNCNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCPWO	NCNCNBW	0.25	0.75	0	0.3	0.33	0.0925	0.0799
SCWCO	NCNCNBW	0.25	0.2	0	0.2	0.16	0.0449	0.0329

SCWBWO	NCNCNBW	0.25	0.2	0	0.2	0.16	0.0449	0.0329
PCO	NCNCNBW	0.2001	0	0	0.2	0.10	0.0269	0.0191
PBWO	NCNCNBW	0.15	0	0	0.2	0.09	0.0237	0.0190
SCCBW	NCNCNBW	0.15	0.2	0	0.2	0.14	0.0384	0.0327
PCBW	NCNCNBW	0.15	0	0	0.2	0.09	0.0237	0.0190
SCPBW	NCNCNBW	0.2	0.2	0	0.2	0.15	0.0416	0.0328
SCPC	NCNCNBW	0.2	1	0	0.2	0.35	0.1007	0.0876
SCPCBW	NCNCNBW	1	0	0	1	0.50	0.1345	0.0955
ALLOK	NCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	NCNCBW	0.4999	0.8	0.85	0.699	0.71	0.1896	0.1944
PO	NCNCBW	0.2001	0	0	0.3	0.13	0.0339	0.0285
SCPWO	NCNCBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWCO	NCNCBW	0.3	0	0	0.4	0.18	0.0474	0.0380
SCWBWO	NCNCBW	0.9	0.1	1	0.4	0.60	0.1515	0.1318
PCO	NCNCBW	0.2001	0	0	0.25	0.11	0.0304	0.0238
PBWO	NCNCBW	0.3	0	0	0.4	0.18	0.0474	0.0380
SCCBW	NCNCBW	0.3	0.1	0	0.4	0.20	0.0547	0.0449
PCBW	NCNCBW	0.25	0	0	0.4	0.16	0.0441	0.0380
SCPBW	NCNCBW	0.3	0.1	0	0.301	0.18	0.0478	0.0356
SCPC	NCNCBW	0.25	0.1	0	0.4	0.19	0.0515	0.0448
SCPCBW	NCNCBW	1	1	0	1	0.75	0.2082	0.1640

ALLOK	NCCPNBW	0	1	0	0	0.25	0.0738	0.0685
SCWO	NCCPNBW	0.25	0.8	0.85	0.101	0.50	0.1316	0.1378
PO	NCCPNBW	0.25	0	0	0.101	0.09	0.0232	0.0099
SCPWO	NCCPNBW	0.25	0	0	0.201	0.11	0.0302	0.0193
SCWCO	NCCPNBW	0.9	0.5	0	0.801	0.55	0.1510	0.1109
SCWBWO	NCCPNBW	0.35	0	1	0.3	0.41	0.1017	0.1147
PCO	NCCPNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PBWO	NCCPNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCCBW	NCCPNBW	0.35	0.5	0	0.201	0.26	0.0735	0.0537
PCBW	NCCPNBW	0.25	0	1	0.2	0.36	0.0883	0.1052
SCPBW	NCCPNBW	0.25	0.15	0	0.3	0.18	0.0482	0.0388
SCPC	NCCPNBW	0.35	0.15	0	0.201	0.18	0.0477	0.0297
SCPCBW	NCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	NCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	NCCPBW	0.25	0.6	0.85	0.201	0.48	0.1239	0.1335
PO	NCCPBW	0.25	0	0	0.201	0.11	0.0302	0.0193
SCPWO	NCCPBW	0.3	0.5	0	0.3	0.28	0.0772	0.0629
SCWCO	NCCPBW	0.35	0.25	1	0.3	0.48	0.1201	0.1319
SCWBWO	NCCPBW	0.35	0.05	0	0.3	0.18	0.0472	0.0321
PCO	NCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PBWO	NCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286

SCCBW	NCCPBW	0.9	0	1	0.85	0.69	0.1757	0.1673
PCBW	NCCPBW	0.35	0.9	0	0.3	0.39	0.1099	0.0904
SCPBW	NCCPBW	0.35	0	1	0.2	0.39	0.0947	0.1053
SCPC	NCCPBW	0.35	0.1	1	0.2	0.41	0.1021	0.1122
SCPCBW	NCCPBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWONCNCNBW	0	0	0.8	0	0.20	0.0465	0.0688
SCWO	IWONCNCNBW	0.25	0	0	0.201	0.11	0.0302	0.0193
PO	IWONCNCNBW	0.9	0.25	1	0.801	0.74	0.1907	0.1798
SCPWO	IWONCNCNBW	0.25	0.2	1	0.2	0.41	0.1030	0.1189
SCWCO	IWONCNCNBW	0.2001	0.5	1	0.2	0.48	0.1219	0.1394
SCWBWO	IWONCNCNBW	0.2001	0	1	0.2	0.35	0.0850	0.1051
PCO	IWONCNCNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PBWO	IWONCNCNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCCBW	IWONCNCNBW	0.2001	0.5	0	0.2	0.23	0.0638	0.0533
PCBW	IWONCNCNBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
SCPBW	IWONCNCNBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
SCPC	IWONCNCNBW	0.25	0.05	0	0.099	0.10	0.0267	0.0131
SCPCBW	IWONCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWONCNCBW	0	0	0.8	0	0.20	0.0465	0.0688
SCWO	IWONCNCBW	0.25	0	0	0.201	0.11	0.0302	0.0193
PO	IWONCNCBW	0.35	0	0.6	0.3	0.31	0.0784	0.0803

SCPWO	IWONCNCBW	0.35	0	0	0.3	0.16	0.0436	0.0287
SCWCO	IWONCNCBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWBWO	IWONCNCBW	0.35	0	0	0.4	0.19	0.0506	0.0381
PCO	IWONCNCBW	0.35	0	0	0.4	0.19	0.0506	0.0381
PBWO	IWONCNCBW	0.9	0	1	0.9	0.70	0.1792	0.1720
SCCBW	IWONCNCBW	0.3	0.5	0	0.3	0.28	0.0772	0.0629
PCBW	IWONCNCBW	0.35	0.6	0	0.3	0.31	0.0878	0.0698
SCPBW	IWONCNCBW	0.35	0.05	0	0.3	0.18	0.0472	0.0321
SCPC	IWONCNCBW	0.3	0.05	0	0.201	0.14	0.0371	0.0228
SCPCBW	IWONCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWONCCPNBW	0	0	0.8	0	0.20	0.0465	0.0688
SCWO	IWONCCPNBW	0.25	0	0	0.101	0.09	0.0232	0.0099
PO	IWONCCPNBW	0.35	0	0.6	0.2	0.29	0.0714	0.0709
SCPWO	IWONCCPNBW	0.35	0	0	0.4	0.19	0.0506	0.0381
SCWCO	IWONCCPNBW	0.35	0	0	0.4	0.19	0.0506	0.0381
SCWBWO	IWONCCPNBW	0.25	0	0	0.201	0.11	0.0302	0.0193
PCO	IWONCCPNBW	0.9	0.1	0	0.9	0.48	0.1284	0.0928
PBWO	IWONCCPNBW	0.35	0	1	0.3	0.41	0.1017	0.1147
SCCBW	IWONCCPNBW	0.25	0.8	0	0.2	0.31	0.0891	0.0740
PCBW	IWONCCPNBW	0.3	0.8	0	0.3	0.35	0.0994	0.0834
SCPBW	IWONCCPNBW	0.25	0.15	0	0.201	0.15	0.0412	0.0295

SCPC	IWONCCPNBW	0.35	0.15	0	0.201	0.18	0.0477	0.0297
SCPCBW	IWONCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWONCCPBW	0	0	0.8	0	0.20	0.0465	0.0688
SCWO	IWONCCPBW	0.25	0	0	0.201	0.11	0.0302	0.0193
PO	IWONCCPBW	0.35	0.1	0.6	0.3	0.34	0.0858	0.0872
SCPWO	IWONCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWCO	IWONCCPBW	0.3	0	0	0.201	0.13	0.0334	0.0193
SCWBWO	IWONCCPBW	0.35	0	0	0.3	0.16	0.0436	0.0287
PCO	IWONCCPBW	0.4	0	0.5	0.3	0.30	0.0758	0.0718
PBWO	IWONCCPBW	0.4	0	0.5	0.4	0.33	0.0829	0.0812
SCCBW	IWONCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PCBW	IWONCCPBW	0.9	0.5	1	0.85	0.81	0.2125	0.2015
SCPBW	IWONCCPBW	0.3	0.2	0	0.201	0.18	0.0481	0.0330
SCPC	IWONCCPBW	0.3	0.2	0	0.201	0.18	0.0481	0.0330
SCPCBW	IWONCCPBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWOCNCNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IWOCNCNBW	0.25	0.8	0	0.099	0.29	0.0820	0.0645
PO	IWOCNCNBW	0.25	0	0.8	0.201	0.31	0.0767	0.0881
SCPWO	IWOCNCNBW	0.9	0.3	1	0.899	0.77	0.2012	0.1924
SCWCO	IWOCNCNBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
SCWBWO	IWOCNCNBW	0.25	0.1	0	0.201	0.14	0.0376	0.0261

PCO	IWOCNCNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PBWO	IWOCNCNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCCBW	IWOCNCNBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
PCBW	IWOCNCNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCPBW	IWOCNCNBW	0.35	0.1	0	0.3	0.19	0.0509	0.0356
SCPC	IWOCNCNBW	0.35	0.3	0	0.2	0.21	0.0587	0.0399
SCPCBW	IWOCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWOCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IWOCNCBW	0.25	0.8	0	0.3	0.34	0.0961	0.0834
PO	IWOCNCBW	0.25	0	0	0.299	0.14	0.0371	0.0285
SCPWO	IWOCNCBW	0.35	0.3	0	0.399	0.26	0.0726	0.0586
SCWCO	IWOCNCBW	0.25	0.1	0	0.201	0.14	0.0376	0.0261
SCWBWO	IWOCNCBW	0.35	0.1	0	0.299	0.19	0.0509	0.0355
PCO	IWOCNCBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PBWO	IWOCNCBW	0.35	0.05	0.85	0.3	0.39	0.0967	0.1053
SCCBW	IWOCNCBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261
PCBW	IWOCNCBW	0.3	0	0	0.2	0.13	0.0333	0.0193
SCPBW	IWOCNCBW	0.9	0.1	1	0.899	0.72	0.1865	0.1787
SCPC	IWOCNCBW	0.3	0.3	0	0.201	0.20	0.0555	0.0399
SCPCBW	IWOCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWOCCPNBW	0	0	0	0	0.00	0.0000	0.0000

SCWO	IWOCCPNBW	0.25	0.8	0	0.2	0.31	0.0891	0.0740
PO	IWOCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCPWO	IWOCCPNBW	0.35	0.3	0	0.3	0.24	0.0657	0.0493
SCWCO	IWOCCPNBW	0.35	0.1	0	0.3	0.19	0.0509	0.0356
SCWBWO	IWOCCPNBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
PCO	IWOCCPNBW	0.35	0.05	0.75	0.201	0.34	0.0839	0.0874
PBWO	IWOCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCCBW	IWOCCPNBW	0.35	0.1	0	0.2	0.16	0.0439	0.0262
PCBW	IWOCCPNBW	0.35	0	0	0.3	0.16	0.0436	0.0287
SCPBW	IWOCCPNBW	0.35	0.1	0	0.2	0.16	0.0439	0.0262
SCPC	IWOCCPNBW	0.9	0.3	1	0.899	0.77	0.2012	0.1924
SCPCBW	IWOCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IWOCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IWOCCPBW	0.2	0.8	0	0.2	0.30	0.0859	0.0739
PO	IWOCCPBW	0.2	0	0	0.201	0.10	0.0270	0.0192
SCPWO	IWOCCPBW	0.25	0.3	0	0.3	0.21	0.0593	0.0491
SCWCO	IWOCCPBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
SCWBWO	IWOCCPBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
PCO	IWOCCPBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
PBWO	IWOCCPBW	0.25	0.05	0	0.25	0.14	0.0373	0.0273
SCCBW	IWOCCPBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355

PCBW	IWOCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCPBW	IWOCCPBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261
SCPC	IWOCCPBW	0.3	0.3	0	0.2	0.20	0.0555	0.0398
SCPCBW	IWOCCPBW	1	1	1	1	1.00	0.2664	0.2500
ALLOK	EWONCNCNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWONCNCNBW	0.3	0	0	0.4	0.18	0.0474	0.0380
PO	EWONCNCNBW	0.9	0.25	1	0.6	0.69	0.1766	0.1609
SCPWO	EWONCNCNBW	0.3	0.2	0	0.3	0.20	0.0551	0.0423
SCWCO	EWONCNCNBW	0.2	0.5	0	0.199	0.22	0.0637	0.0533
SCWBWO	EWONCNCNBW	0.2	0	0	0.3	0.13	0.0339	0.0285
PCO	EWONCNCNBW	0.2	0	0	0.3	0.13	0.0339	0.0285
PBWO	EWONCNCNBW	0.2	0	0	0.199	0.10	0.0268	0.0190
SCCBW	EWONCNCNBW	0.2	0.5	0	0.199	0.22	0.0637	0.0533
PCBW	EWONCNCNBW	0.2	0.05	0	0.199	0.11	0.0305	0.0224
SCPBW	EWONCNCNBW	0.2	0.05	0	0.3	0.14	0.0376	0.0319
SCPC	EWONCNCNBW	0.2	0.05	0	0.199	0.11	0.0305	0.0224
SCPCBW	EWONCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWONCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWONCNCBW	0.3	0	0	0.2	0.13	0.0333	0.0193
PO	EWONCNCBW	0.3	0	0.75	0.4	0.36	0.0910	0.1026
SCPWO	EWONCNCBW	0.3	0	0	0.201	0.13	0.0334	0.0193

SCWCO	EWONCNCBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCWBWO	EWONCNCBW	0.2	0	0	0.3	0.13	0.0339	0.0285
CO	EWONCNCBW	0.25	0	0	0.201	0.11	0.0302	0.0193
PBWO	EWONCNCBW	0.9	0	1	0.85	0.69	0.1757	0.1673
SCCBW	EWONCNCBW	0.25	0.5	0	0.201	0.24	0.0671	0.0535
PCBW	EWONCNCBW	0.3	0.6	0	0.3	0.30	0.0846	0.0697
SCPBW	EWONCNCBW	0.3	0.05	0	0.2	0.14	0.0370	0.0227
CPC	EWONCNCBW	0.3	0.05	0	0.3	0.16	0.0440	0.0321
SCPCBW	EWONCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWONCCPNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWONCCPNBW	0.3	0	0	0.2	0.13	0.0333	0.0193
PO	EWONCCPNBW	0.3	0	0	0.2	0.13	0.0333	0.0193
SCPWO	EWONCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCWCO	EWONCCPNBW	0.25	0	0	0.6	0.21	0.0582	0.0568
SCWBWO	EWONCCPNBW	0.2	0	0	0.3	0.13	0.0339	0.0285
PCO	EWONCCPNBW	0.9	0	1	0.8	0.68	0.1722	0.1626
PBWO	EWONCCPNBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
SCCBW	EWONCCPNBW	0.2	0	0	0.4	0.15	0.0409	0.0379
PCBW	EWONCCPNBW	0.25	0.8	0	0.301	0.34	0.0962	0.0835
SCPBW	EWONCCPNBW	0.2	0.8	0	0.3	0.33	0.0929	0.0833
SCPC	EWONCCPNBW	0.3	0.15	0	0.2	0.16	0.0444	0.0295

SCPCBW	EWONCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWONCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWONCCPBW	0.25	0	0	0.2	0.11	0.0301	0.0192
PO	EWONCCPBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355
SCPWO	EWONCCPBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCWCO	EWONCCPBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCWBWO	EWONCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PCO	EWONCCPBW	0.35	0	0	0.4	0.19	0.0506	0.0381
PBWO	EWONCCPBW	0.35	0	0	0.201	0.14	0.0366	0.0194
SCCBW	EWONCCPBW	0.25	0	0	0.2	0.11	0.0301	0.0192
PCBW	EWONCCPBW	0.9	0.5	1	0.899	0.82	0.2160	0.2061
SCPBW	EWONCCPBW	0.25	0.2	0	0.201	0.16	0.0449	0.0330
SCPC	EWONCCPBW	0.25	0.2	0	0.201	0.16	0.0449	0.0330
SCPCBW	EWONCCPBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWOCNCNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWOCNCNBW	0.25	0.8	0	0.1	0.29	0.0821	0.0646
PO	EWOCNCNBW	0.25	0	0.301	0.1	0.16	0.0406	0.0357
SCPWO	EWOCNCNBW	0.9	0.3	1	0.901	0.78	0.2014	0.1926
SCWCO	EWOCNCNBW	0.2	0.1	0	0.2	0.13	0.0343	0.0259
SCWBWO	EWOCNCNBW	0.2	0.1	0	0.2	0.13	0.0343	0.0259
PCO	EWOCNCNBW	0.2	0	0	0.3	0.13	0.0339	0.0285

PBWO	EWOCNCNBW	0.2	0	0	0.2	0.10	0.0269	0.0191
SCCBW	EWOCNCNBW	0.2	0.1	0	0.2	0.13	0.0343	0.0259
PCBW	EWOCNCNBW	0.2	0	0	0.3	0.13	0.0339	0.0285
SCPBW	EWOCNCNBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
SCPC	EWOCNCNBW	0.25	0.3	0	0.2	0.19	0.0522	0.0397
SCPCBW	EWOCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWOCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWOCNCBW	0.2001	0.8	0	0.2	0.30	0.0859	0.0739
PO	EWOCNCBW	0.2001	0	0	0.3	0.13	0.0339	0.0285
SCPWO	EWOCNCBW	0.3	0.3	0	0.3	0.23	0.0625	0.0492
SCWCO	EWOCNCBW	0.2001	0.1	0	0.4	0.18	0.0483	0.0447
SCWBWO	EWOCNCBW	0.3	0.1	0	0.4	0.20	0.0547	0.0449
PCO	EWOCNCBW	0.2001	0	0	0.201	0.10	0.0270	0.0192
PBWO	EWOCNCBW	0.3	0.05	0.35	0.3	0.25	0.0644	0.0622
SCCBW	EWOCNCBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
PCBW	EWOCNCBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCPBW	EWOCNCBW	0.9	0.1	1	0.9	0.73	0.1866	0.1788
SCPC	EWOCNCBW	0.25	0.3	0	0.3	0.21	0.0593	0.0491
SCPCBW	EWOCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWOCCPNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWOCCPNBW	0.2001	0.8	0	0.201	0.30	0.0860	0.0740

PO	EWOCCPNBW	0.2001	0	0	0.201	0.10	0.0270	0.0192
SCPWO	EWOCCPNBW	0.3	0.3	0	0.2	0.20	0.0555	0.0398
SCWCO	EWOCCPNBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355
SCWBWO	EWOCCPNBW	0.2001	0.1	0	0.201	0.13	0.0343	0.0260
PCO	EWOCCPNBW	0.2001	0.05	0.3501	0.201	0.20	0.0510	0.0527
PBWO	EWOCCPNBW	0.2001	0	0	0.1	0.08	0.0199	0.0097
SCCBW	EWOCCPNBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355
PCBW	EWOCCPNBW	0.3	0	0	0.2	0.13	0.0333	0.0193
SCPBW	EWOCCPNBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355
SCPC	EWOCCPNBW	0.9	0.3	1	0.9	0.78	0.2013	0.1925
SCPCBW	EWOCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	EWOCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	EWOCCPBW	0.15	0.8	0	0.2	0.29	0.0827	0.0738
PO	EWOCCPBW	0.15	0	0	0.2	0.09	0.0237	0.0190
SCPWO	EWOCCPBW	0.2	0.3	0	0.3	0.20	0.0560	0.0490
SCWCO	EWOCCPBW	0.2	0.1	0	0.3	0.15	0.0413	0.0353
SCWBWO	EWOCCPBW	0.2	0.1	0	0.3	0.15	0.0413	0.0353
PCO	EWOCCPBW	0.2	0.05	0	0.4	0.16	0.0446	0.0413
PBWO	EWOCCPBW	0.2	0.05	0	0.3	0.14	0.0376	0.0319
SCCBW	EWOCCPBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
PCBW	EWOCCPBW	0.25	0	0	0.4	0.16	0.0441	0.0380

SCPBW	EWOCCPBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
SCPC	EWOCCPBW	0.25	0.3	0	0.201	0.19	0.0523	0.0398
SCPCBW	EWOCCPBW	1	1	1	1	1.00	0.2664	0.2500
ALLOK	IEWNCNCNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWNCNCNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PO	IEWNCNCNBW	0.9	0.25	1	0.8	0.74	0.1906	0.1797
SCPWO	IEWNCNCNBW	0.35	0.2	0	0.3	0.21	0.0583	0.0424
SCWCO	IEWNCNCNBW	0.25	0.5	0	0.2	0.24	0.0670	0.0534
SCWBWO	IEWNCNCNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
PCO	IEWNCNCNBW	0.25	0	0	0.301	0.14	0.0372	0.0287
PBWO	IEWNCNCNBW	0.25	0	0	0.1	0.09	0.0231	0.0098
SCCBW	IEWNCNCNBW	0.25	0.5	0	0.2	0.24	0.0670	0.0534
PCBW	IEWNCNCNBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
SCPBW	IEWNCNCNBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
SCPC	IEWNCNCNBW	0.25	0.05	0	0.2	0.13	0.0338	0.0226
SCPCBW	IEWNCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IEWNCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWNCNCBW	0.3	0	0	0.201	0.13	0.0334	0.0193
PO	IEWNCNCBW	0.35	0	0	0.2	0.14	0.0365	0.0193
SCPWO	IEWNCNCBW	0.35	0	0	0.2	0.14	0.0365	0.0193
CWCO	IEWNCNCBW	0.3	0	0	0.3	0.15	0.0403	0.0286

SCWBWO	IEWNCNCBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PCO	IEWNCNCBW	0.3	0	0	0.3	0.15	0.0403	0.0286
PBWO	IEWNCNCBW	0.9	0	1	0.9	0.70	0.1792	0.1720
SCCBW	IEWNCNCBW	0.3	0.5	0	0.3	0.28	0.0772	0.0629
PCBW	IEWNCNCBW	0.35	0.6	0	0.4	0.34	0.0948	0.0792
CPBW	IEWNCNCBW	0.35	0.05	0	0.3	0.18	0.0472	0.0321
SCPC	IEWNCNCBW	0.35	0.05	0	0.3	0.18	0.0472	0.0321
SCPCBW	IEWNCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IEWNCCPNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWNCCPNBW	0.3	0	0	0.2	0.13	0.0333	0.0193
PO	IEWNCCPNBW	0.35	0	0	0.3	0.16	0.0436	0.0287
SCPWO	IEWNCCPNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWCO	IEWNCCPNBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWBWO	IEWNCCPNBW	0.25	0	0	0.4	0.16	0.0441	0.0380
PCO	IEWNCCPNBW	0.9	0.1	1	0.8	0.70	0.1795	0.1694
PBWO	IEWNCCPNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
SCCBW	IEWNCCPNBW	0.25	0.8	0	0.3	0.34	0.0961	0.0834
PCBW	IEWNCCPNBW	0.3	0.8	0	0.4	0.38	0.1064	0.0928
SCPBW	IEWNCCPNBW	0.25	0.15	0	0.2	0.15	0.0412	0.0294
SCPC	IEWNCCPNBW	0.3	0.15	0	0.2	0.16	0.0444	0.0295
SCPCBW	IEWNCCPNBW	1	1	0	1	0.75	0.2082	0.1640

ALLOK	IEWNCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWNCCPBW	0.25	0	0	0.2	0.11	0.0301	0.0192
PO	IEWNCCPBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261
SCPWO	IEWNCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCWCO	IEWNCCPBW	0.3	0	0	0.4	0.18	0.0474	0.0380
SCWBWO	IEWNCCPBW	0.3	0	0	0.4	0.18	0.0474	0.0380
PCO	IEWNCCPBW	0.35	0	0	0.3	0.16	0.0436	0.0287
PBWO	IEWNCCPBW	0.35	0	0	0.3	0.16	0.0436	0.0287
SCCBW	IEWNCCPBW	0.3	0	1	0.201	0.38	0.0915	0.1054
PCBW	IEWNCCPBW	0.9	0.5	0	0.801	0.55	0.1510	0.1109
SCPBW	IEWNCCPBW	0.3	0.2	0	0.3	0.20	0.0551	0.0423
SCPC	IEWNCCPBW	0.3	0.2	0	0.4	0.23	0.0621	0.0517
SCPCBW	IEWNCCPBW	1	1	0	1	0.75	0.2082	0.1640
LLOK	IEWCNCNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWCNCNBW	0.25	0.8	0	0.3	0.34	0.0961	0.0834
PO	IEWCNCNBW	0.25	0	1	0.3	0.39	0.0953	0.1146
SCPWO	IEWCNCNBW	0.9	0.3	0	0.85	0.51	0.1397	0.1018
SCWCO	IEWCNCNBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
SCWBWO	IEWCNCNBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
PCO	IEWCNCNBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PBWO	IEWCNCNBW	0.25	0	0	0.2	0.11	0.0301	0.0192

SCCBW	IEWCNCNBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
PCBW	IEWCNCNBW	0.25	0	0	0.4	0.16	0.0441	0.0380
CPBW	IEWCNCNBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261
SCPC	IEWCNCNBW	0.3	0.3	0	0.2	0.20	0.0555	0.0398
SCPCBW	IEWCNCNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IEWCNCBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWCNCBW	0.25	0.8	0	0.3	0.34	0.0961	0.0834
PO	IEWCNCBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCPWO	IEWCNCBW	0.35	0.3	0	0.3	0.24	0.0657	0.0493
SCWCO	IEWCNCBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
SCWBWO	IEWCNCBW	0.35	0.1	0	0.301	0.19	0.0510	0.0357
PCO	IEWCNCBW	0.25	0	0	0.3	0.14	0.0371	0.0286
PBWO	IEWCNCBW	0.35	0.05	0	0.201	0.15	0.0403	0.0228
SCCBW	IEWCNCBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261
PCBW	IEWCNCBW	0.3	0	0	0.2	0.13	0.0333	0.0193
SCPBW	IEWCNCBW	0.9	0.1	1	0.85	0.71	0.1830	0.1741
SCPC	IEWCNCBW	0.3	0.3	0	0.2	0.20	0.0555	0.0398
SCPCBW	IEWCNCBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IEWCCPNBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWCCPNBW	0.25	0.8	0	0.2	0.31	0.0891	0.0740
PO	IEWCCPNBW	0.25	0	0	0.299	0.14	0.0371	0.0285

SCPWO	IEWCCPNBW	0.35	0.3	0	0.3	0.24	0.0657	0.0493
SCWCO	IEWCCPNBW	0.35	0.1	0	0.3	0.19	0.0509	0.0356
SCWBWO	IEWCCPNBW	0.25	0.1	0	0.3	0.16	0.0445	0.0354
PCO	IEWCCPNBW	0.35	0.05	0	0.3	0.18	0.0472	0.0321
PBWO	IEWCCPNBW	0.25	0	0	0.2	0.11	0.0301	0.0192
SCCBW	IEWCCPNBW	0.35	0.1	0	0.3	0.19	0.0509	0.0356
PCBW	IEWCCPNBW	0.35	0	0	0.3	0.16	0.0436	0.0287
SCPBW	IEWCCPNBW	0	0.1	0	0.499	0.15	0.0424	0.0537
SCPC	IEWCCPNBW	0.9	0.3	1	0.8	0.75	0.1943	0.1831
SCPCBW	IEWCCPNBW	1	1	0	1	0.75	0.2082	0.1640
ALLOK	IEWCCPBW	0	0	0	0	0.00	0.0000	0.0000
SCWO	IEWCCPBW	0.2	0.8	0	0.2	0.30	0.0859	0.0739
PO	IEWCCPBW	0.2	0	0	0.2	0.10	0.0269	0.0191
SCPWO	IEWCCPBW	0.25	0.3	0	0.3	0.21	0.0593	0.0491
SCWCO	IEWCCPBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
SCWBWO	IEWCCPBW	0.25	0.1	0	0.2	0.14	0.0375	0.0260
PCO	IEWCCPBW	0.25	0.05	0	0.3	0.15	0.0408	0.0320
PBWO	IEWCCPBW	0.25	0.05	0	0.201	0.13	0.0339	0.0227
SCCBW	IEWCCPBW	0.3	0.1	0	0.3	0.18	0.0477	0.0355
PCBW	IEWCCPBW	0.3	0	0	0.3	0.15	0.0403	0.0286
SCPBW	IEWCCPBW	0.3	0.1	0	0.2	0.15	0.0407	0.0261

SCPC	IEWCCPBW	0.3	0.3	0	0.2	0.20	0.0555	0.0398
SCPCBW	IEWCCPBW	1	0	0	1	0.50	0.1345	0.0955

A.4 Aggregated expert assessments for load test outcome spreading tendon inspection test

The table below compares the aggregated assessments of the experts on their judgements on load test outcome spreading tendon test. The comparison is between the nonparametric methods, EB methods and simple average. The weights used for the experts for EB aggregation are : E1=0.2174177, E2=0.2961574, E3=0.2070601 and E4=0.2793648. The weights used for the experts for Non-parametric aggregation are: E1=0.199832666, E2=0.503325469, E3=0.007008922 and E4=0.289832943.

Outcome	Condition	E1	E2	E3	E4	Average	EB	NP
FE	OLAS	1	1	1	0	0.75	0.1801588	0.177541764
TF	OLAS	0	1	0	0	0.25	0.07403935	0.125831367
FE	OLNAS	0.95	1	1	0.95	0.975	0.243790219	0.24387918
TF	OLNAS	1	0	0	0	0.25	0.054354425	0.049958167
FE	RCALAS	1	1	1	0.9	0.975	0.24301588	0.242754176
TF	RCALAS	0	1	0	0	0.25	0.07403935	0.125831367
FE	RCALNAS	1	1	1	1	1	0.25	0.25
TF	RCALNAS	0	1	0	0	0.25	0.07403935	0.125831367

FE	RCUALAS	0.9	0.85	0.9	0.6	0.8125	0.200345673	0.196970961
TF	RCUALAS	0	0.99	0	0	0.2475	0.073298957	0.124573054
FE	RCUALNAS	0.9	0.85	0.9	0.3	0.7375	0.179393313	0.17523349
TF	RCUALNAS	0	1	1	0	0.5	0.125804375	0.127583598
FE	FAS	0	0	0.3	0	0.075	0.015529508	0.000525669
TF	FAS	0	1	0	0	0.25	0.07403935	0.125831367
FE	FNAS	0	0	0.3	0	0.075	0.015529508	0.000525669
TF	FNAS	0	1	0	0	0.25	0.07403935	0.125831367

A.5 Aggregated expert assessments for lab tensile test on strand

The table below compares the aggregated assessments of the experts on their judgements on load test outcome spreading tendon test. The comparison is between the nonparametric methods, EB methods and simple average. The weights used for the experts for EB aggregation are : E1=0.3311689, E2=0.2149394, E3=0.2458123 and E4=0.2080794. The weights used for the experts for Non-parametric aggregation are: E1=0.57888202, E2=0.10879146, E3=0.22915698 and E4=0.08316955.

Outcome	Condition	E1	E2	E3	E4	Average	EB	NP
FWEEL	OL	0.99	1	0.95	1	0.985	0.2461	0.2457
FWNEEL	OL	1	0	1	0	0.5	0.1442	0.2020

FWEEL	RCAL	0.8	0.95	0.9499	0.9	0.899975	0.2225	0.2147
FWNEEL	RCAL	1	1	1	1	1	0.2500	0.2500
FWEEL	RCUAL	0.6	0.25	0.85	0.7	0.6	0.1518	0.1569
FWNEEL	RCUAL	1	1	0	0	0.5	0.1365	0.1719
FWEEL	F	0	0	0.0501	0	0.012525	0.0031	0.0029
FWNEEL	F	1	1	1	1	1	0.2500	0.2500

Appendix B

Model derivations for aggregation of correlated judgements with bias

This chapter presents two models for dealing with biases in the correlated judgements. Section B1 presents the non-parametric model with bias and section B2 presents the EB parametric model with bias. These models have not been implemented in the main chapters as the models have been developed in absence of complete truth around the parameters of interest. Therefore, adding a bias term to it would add to the systematic errors in both the models. However, for the sake of completeness of the models, both the theorems dealing with biases are presented here in the appendix.

B.1 Non-parametric aggregation model with bias

Theorem 3 *Let the mean square error be defined as* $MSE = E\left((\sum w_i X_i - \mu b_i)^2\right)$, *for* $k = 1$

where b_i denotes the multiplicative bias that exists in the judgements and C denotes the covariance matrix; then the optimal weights in terms of Mean Square Error, for aggregating the judgements are proportional to the solution of $\underline{\hat{w}}_{b_{ki}} \propto C^{-1}$, where, $\underline{\hat{w}}_{b_{ki}} = \left(\frac{\hat{w}_1}{b_1}, ..., \frac{\hat{w}_n}{b_n}\right)$, is the vector of weights applied to the assessments of experts.

Proof 3 $\underline{\hat{w}} = argminS_{\forall w_i} = argmin\forall w_i E\left(\left(\sum_{\forall i} w_i X_i - \mu\right)^2\right)$

Using the Lagrange multiplier approach,

$L = E\left(\left(\sum_{\forall i} w_i X_i - \mu\right)^2\right) + \lambda\left(\sum_{i=1}^{n} w_i - 1\right)$

The first order conditions with respect to the weights are as in the following for which there will be n such equations, one for each i,

$\frac{dL}{dw_i} = E\left(2\left(\sum_{\forall i} w_i X_i - \mu\right) X_i\right) + \lambda$

This is then set to zero and results in the following with $\widehat{w_i}$ used to represent the estimated weight.

$\sum_{\forall i} \hat{w}_i E\left(X_i X_i\right) - \mu E\left(X_i\right) = \lambda$

Given the assumption of biasedness, for each i,

$\sum_{\forall i} \hat{w}_i \left(c_{ij} + (b_i \mu)^2\right) = \lambda$

The first order condition with respect to the constraint is as in the following,

$\frac{dL}{d\lambda} = \sum_{i=1}^{n} \hat{w}_i - 1 = 0$

Substituting back into the first order constraints with respect to the weights, the following equation is thus obtained,

$\sum_{\forall j} \hat{w}_i c_{ij} = \lambda$

Therefore,

$\underline{\hat{w}}_b \propto C^{-1}$

Lemma 1 *Assume a sample of n unbiased experts each providing a prediction, X_{ik}, where $i = 1, 2, ..., n$ and $k = 1, 2, ..., m$ on some unknown quantity θ_k. If the moments are denoted as $E(X_{ik}) = \mu_k + b_{ik}$ and $C(X_i X_j) = c_{ij}$, where b_{ik} denotes the additive bias that exists in the judgements and C denotes the covariance matrix; then the optimal weights in terms of Mean Square Error (S), for aggregating the judgements are proportional to the solution of $\underline{\hat{w}}_b \propto C^{-1}$, where, $\underline{\hat{w}}_b = (\underline{\hat{w}}_1, ..., \underline{\hat{w}}_n)$, is the vector of weights applied to the assessments of experts.*

B.2 EB model with bias

Assume that an expert provides us his estimate x_1 and the difference between the estimate and the true value is represented by a location bias error, e, such that, $x_1 = \mu_1 + e_1$. The workings are assuming a normal- normal model (as described in chapter four of this thesis) adjusted for location bias. The model presented here is assuming four experts and the judgements are treated as variables, $x_1, ..., x_4$, then consider the judgements as sum of two terms, $x_1 = \mu_1 + e_1, x_2 = \mu_2 + e_2, x_3 = \mu_3 + e_3, x_4 = \mu_4 + e_4$ where e represents the error terms; then on taking expectations,

$$x_1 = \mu + b_1$$

$$x_2 = \mu + b_2$$

$$x_3 = \mu + b_3$$

$$x_4 = \mu + b_4$$

where

$$E\left(e_1\right)=b_1$$

and b, is the bias.

The above equations show that the expert judgements are assumed to be biased by the additive quantities of

$$b_1,b_2,b_3,b_4$$

. The likelihood function as a multivariate normal for four experts,

$$L\left(\mu|x_1,x_2,x_3,x_4\right)=\exp\left(\begin{array}{c}-\frac{1}{2(1-\rho^2)}\left(\begin{array}{c}\frac{(x_1-(\mu+b_1))^2}{\sigma^2{}_1}+\frac{(x_2-(\mu+b_2))^2}{\sigma^2{}_2}+\\ \frac{(x_3-(\mu+b_3))^2}{\sigma^2{}_3}+\frac{(x_4-(\mu+b_4))^2}{\sigma^2{}_4}\end{array}\right)-\\ \left(\begin{array}{c}\frac{2\rho_{12}(x_1-(\mu+b_1))(x_2-(\mu+b_2))}{\sigma_1\sigma_2}+\frac{2\rho_{13}(x_1-(\mu+b_1))(x_3-(\mu+b_3))}{\sigma_1\sigma_3}\\ +\frac{2\rho_{14}(x_1-(\mu+b_1))(x_4-(\mu+b_4))}{\sigma_1\sigma_4}+\frac{2\rho_{23}(x_2-(\mu+b_2))(x_3-(\mu+b_3))}{\sigma_2\sigma_3}\\ +\frac{2\rho_{24}(x_2-(\mu+b_2))(x_4-(\mu+b_4))}{\sigma_2\sigma_4}+\frac{2\rho_{34}(x_3-(\mu+b_3))(x_4-(\mu+b_4))}{\sigma_3\sigma_4}\end{array}\right)\end{array}\right)$$

On completing squares, and multiplying with prior, the posterior mean and variance are given as follows,

$$\mu_1 = \frac{\begin{array}{l}\left(\frac{x_1-b_1}{(1-\rho^2){\sigma^2}_1}\right)+\left(\frac{x_2-b_2}{(1-\rho^2){\sigma^2}_2}\right)+\left(\frac{x_3-b_3}{(1-\rho^2){\sigma^2}_3}\right)+\left(\frac{x_4-b_4}{(1-\rho^2){\sigma^2}_4}\right)\\ -\frac{\rho_{12}}{(1-\rho^2)\sigma_1\sigma_2}\left(x_1+x_2+b_1+b_2\right)-\frac{\rho_{13}}{(1-\rho^2)\sigma_1\sigma_3}\left(x_1+x_3+b_1+b_3\right)\\ -\frac{\rho_{14}}{(1-\rho^2)\sigma_1\sigma_4}\left(x_1+x_4+b_1+b_4\right)-\frac{\rho_{23}}{(1-\rho^2)\sigma_2\sigma_3}\left(x_2+x_3+b_2+b_3\right)\\ -\frac{\rho_{24}}{(1-\rho^2)\sigma_2\sigma_4}\left(x_2+x_4+b_2+b_4\right)-\frac{\rho_{34}}{(1-\rho^2)\sigma_3\sigma_4}\left(x_3+x_4+b_3+b_4\right)+\frac{\theta}{\tau^2}\end{array}}{\left(\begin{array}{l}\frac{1}{(1-\rho^2){\sigma^2}_1}+\frac{1}{(1-\rho^2){\sigma^2}_2}+\frac{1}{(1-\rho^2){\sigma^2}_3}+\frac{1}{(1-\rho^2){\sigma^2}_4}-\frac{2\rho_{12}}{(1-\rho^2)\sigma_1\sigma_2}\\ -\frac{2\rho_{13}}{(1-\rho^2)\sigma_1\sigma_3}-\frac{2\rho_{14}}{(1-\rho^2)\sigma_1\sigma_4}-\frac{2\rho_{23}}{(1-\rho^2)\sigma_2\sigma_3}-\frac{2\rho_{24}}{(1-\rho^2)\sigma_2\sigma_4}-\frac{2\rho_{34}}{(1-\rho^2)\sigma_3\sigma_4}+\frac{1}{\tau^2}\end{array}\right)}$$

$$\sigma^2 = \left(\begin{array}{l}\frac{1}{(1-\rho^2){\sigma^2}_1}+\frac{1}{(1-\rho^2){\sigma^2}_2}+\frac{1}{(1-\rho^2){\sigma^2}_3}+\frac{1}{(1-\rho^2){\sigma^2}_4}-\frac{2\rho_{12}}{(1-\rho^2)\sigma_1\sigma_2}\\ -\frac{2\rho_{13}}{(1-\rho^2)\sigma_1\sigma_3}-\frac{2\rho_{14}}{(1-\rho^2)\sigma_1\sigma_4}-\frac{2\rho_{23}}{(1-\rho^2)\sigma_2\sigma_3}-\frac{2\rho_{24}}{(1-\rho^2)\sigma_2\sigma_4}-\frac{2\rho_{34}}{(1-\rho^2)\sigma_3\sigma_4}+\frac{1}{\tau^2}\end{array}\right)$$

It can be seen that the posterior mean is a weighted average of the prior mean and adjusted to the location bias. The workings are below:

$$= \frac{\left({x_1}^2+\mu^2+{b_1}^2-2\mu x_1+2\mu b_1-2b_1x_1\right)}{{\sigma^2}_1}+\frac{\left({x_2}^2+\mu^2+{b_2}^2-2\mu x_2+2\mu b_2-2b_2x_2\right)}{{\sigma^2}_2}+$$
$$\frac{\left({x_3}^2-\mu^2-{b_3}^2-2\mu x_3+2\mu b_3-2b_3x_3\right)}{{\sigma^2}_3}+\frac{\left({x_4}^2-\mu^2-{b_4}^2-2\mu x_4+2\mu b_4-2b_4x_4\right)}{{\sigma^2}_4}$$
$$= \left(\frac{{x_1}^2}{{\sigma^2}_1}+\frac{{x_2}^2}{{\sigma^2}_2}+\frac{{x_3}^2}{{\sigma^2}_3}+\frac{{x_4}^2}{{\sigma^2}_4}\right)+\mu^2\left(\frac{1}{{\sigma^2}_1}+\frac{1}{{\sigma^2}_2}+\frac{1}{{\sigma^2}_3}+\frac{1}{{\sigma^2}_4}\right)+\left(\frac{{b_1}^2}{{\sigma^2}_1}+\frac{{b_2}^2}{{\sigma^2}_2}+\frac{{b_3}^2}{{\sigma^2}_3}+\frac{{b_4}^2}{{\sigma^2}_4}\right)$$
$$-2\mu\left(\left(\frac{x_1-b_1}{{\sigma^2}_1}\right)+\left(\frac{x_2-b_2}{{\sigma^2}_2}\right)+\left(\frac{x_3-b_3}{{\sigma^2}_3}\right)+\left(\frac{x_4-b_4}{{\sigma^2}_4}\right)\right)$$

$$
\mu^2\left(\frac{1}{\sigma^2{}_1}+\frac{1}{\sigma^2{}_2}+\frac{1}{\sigma^2{}_3}+\frac{1}{\sigma^2{}_4}-\frac{2\rho_{12}}{\sigma_1\sigma_2}-\frac{2\rho_{13}}{\sigma_1\sigma_3}-\frac{2\rho_{14}}{\sigma_1\sigma_4}-\frac{2\rho_{23}}{\sigma_2\sigma_3}-\frac{2\rho_{24}}{\sigma_2\sigma_4}-\frac{2\rho_{34}}{\sigma_3\sigma_4}+\frac{1}{\tau^2}\right)
$$

$$
-2\mu\left(\begin{array}{l}
\left(\frac{x_1-b_1}{\sigma^2{}_1}\right)+\left(\frac{x_2-b_2}{\sigma^2{}_2}\right)+\left(\frac{x_3-b_3}{\sigma^2{}_3}\right)+\left(\frac{x_4-b_4}{\sigma^2{}_4}\right)-\frac{\rho_{12}}{\sigma_1\sigma_2}\left(x_1+x_2+b_1+b_2\right)\\
-\frac{\rho_{13}}{\sigma_1\sigma_3}\left(x_1+x_3+b_1+b_3\right)-\frac{\rho_{14}}{\sigma_1\sigma_4}\left(x_1+x_4+b_1+b_4\right)-\frac{\rho_{23}}{\sigma_2\sigma_3}\left(x_2+x_3+b_2+b_3\right)\\
-\frac{\rho_{24}}{\sigma_2\sigma_4}\left(x_2+x_4+b_2+b_4\right)-\frac{\rho_{34}}{\sigma_3\sigma_4}\left(x_3+x_4+b_3+b_4\right)+\frac{\theta}{\tau^2}
\end{array}\right)+
$$

$$
\frac{\theta^2}{\tau^2}+\left(\frac{x_1{}^2}{\sigma^2{}_1}+\frac{x_2{}^2}{\sigma^2{}_2}+\frac{x_3{}^2}{\sigma^2{}_3}+\frac{x_4{}^2}{\sigma^2{}_4}\right)+\left(\frac{b_1{}^2}{\sigma^2{}_1}+\frac{b_2{}^2}{\sigma^2{}_2}+\frac{b_3{}^2}{\sigma^2{}_3}+\frac{b_4{}^2}{\sigma^2{}_4}\right)
$$

$$
-\left(\begin{array}{l}
\frac{2\rho_{12}}{\sigma_1\sigma_2}\left(x_1x_2-x_1b_2+x_2b_1\right)+\frac{2\rho_{13}}{\sigma_1\sigma_3}\left(x_1x_3-x_1b_3+x_3b_1\right)+\frac{2\rho_{14}}{\sigma_1\sigma_4}\left(x_1x_4-x_1b_4+x_4b_1\right)+\\
\frac{2\rho_{23}}{\sigma_2\sigma_3}\left(x_2x_3-x_2b_3+x_3b_2\right)+\frac{2\rho_{24}}{\sigma_2\sigma_4}\left(x_2x_4-x_2b_4+x_4b_2\right)+\frac{2\rho_{34}}{\sigma_3\sigma_4}\left(x_3x_4-x_3b_4+x_4b_3\right)
\end{array}\right)
$$

$$
-\left(\frac{2\rho_{12}b_1b_2}{\sigma_1\sigma_2}+\frac{2\rho_{13}b_1b_3}{\sigma_1\sigma_3}+\frac{2\rho_{14}b_1b_4}{\sigma_1\sigma_4}+\frac{2\rho_{23}b_2b_3}{\sigma_2\sigma_3}+\frac{2\rho_{24}b_2b_4}{\sigma_2\sigma_4}+\frac{2\rho_{34}b_3b_4}{\sigma_3\sigma_4}\right)
$$

On simplifying and substituting the constant terms with c,

$$\mu^2\left(\frac{1}{{\sigma^2}_1}+\frac{1}{{\sigma^2}_2}+\frac{1}{{\sigma^2}_3}+\frac{1}{{\sigma^2}_4}-\frac{2\rho_{12}}{\sigma_1\sigma_2}-\frac{2\rho_{13}}{\sigma_1\sigma_3}-\frac{2\rho_{14}}{\sigma_1\sigma_4}-\frac{2\rho_{23}}{\sigma_2\sigma_3}-\frac{2\rho_{24}}{\sigma_2\sigma_4}-\frac{2\rho_{34}}{\sigma_3\sigma_4}+\frac{1}{\tau^2}\right)$$

$$-2\mu\left(\begin{array}{l}\left(\frac{x_1-b_1}{{\sigma^2}_1}\right)+\left(\frac{x_2-b_2}{{\sigma^2}_2}\right)+\left(\frac{x_3-b_3}{{\sigma^2}_3}\right)+\left(\frac{x_4-b_4}{{\sigma^2}_4}\right)-\frac{\rho_{12}}{\sigma_1\sigma_2}\left(x_1+x_2+b_1+b_2\right)\\ -\frac{\rho_{13}}{\sigma_1\sigma_3}\left(x_1+x_3+b_1+b_3\right)-\frac{\rho_{14}}{\sigma_1\sigma_4}\left(x_1+x_4+b_1+b_4\right)-\frac{\rho_{23}}{\sigma_2\sigma_3}\left(x_2+x_3+b_2+b_3\right)\\ -\frac{\rho_{24}}{\sigma_2\sigma_4}\left(x_2+x_4+b_2+b_4\right)-\frac{\rho_{34}}{\sigma_3\sigma_4}\left(x_3+x_4+b_3+b_4\right)+\frac{\theta}{\tau^2}\end{array}\right)+$$

$$c$$

$$=\left(\frac{\mu-\left(\frac{\begin{array}{l}\left(\frac{x_1-b_1}{{\sigma^2}_1}\right)+\left(\frac{x_2-b_2}{{\sigma^2}_2}\right)+\left(\frac{x_3-b_3}{{\sigma^2}_3}\right)+\left(\frac{x_4-b_4}{{\sigma^2}_4}\right)-\frac{\rho_{12}}{\sigma_1\sigma_2}\left(x_1+x_2+b_1+b_2\right)\\ -\frac{\rho_{13}}{\sigma_1\sigma_3}\left(x_1+x_3+b_1+b_3\right)-\frac{\rho_{14}}{\sigma_1\sigma_4}\left(x_1+x_4+b_1+b_4\right)-\frac{\rho_{23}}{\sigma_2\sigma_3}\left(x_2+x_3+b_2+b_3\right)\\ -\frac{\rho_{24}}{\sigma_2\sigma_4}\left(x_2+x_4+b_2+b_4\right)-\frac{\rho_{34}}{\sigma_3\sigma_4}\left(x_3+x_4+b_3+b_4\right)+\frac{\theta}{\tau^2}\end{array}}{\left(\frac{1}{{\sigma^2}_1}+\frac{1}{{\sigma^2}_2}+\frac{1}{{\sigma^2}_3}+\frac{1}{{\sigma^2}_4}-\frac{2\rho_{12}}{\sigma_1\sigma_2}-\frac{2\rho_{13}}{\sigma_1\sigma_3}-\frac{2\rho_{14}}{\sigma_1\sigma_4}-\frac{2\rho_{23}}{\sigma_2\sigma_3}-\frac{2\rho_{24}}{\sigma_2\sigma_4}-\frac{2\rho_{34}}{\sigma_3\sigma_4}+\frac{1}{\tau^2}\right)}\right)}{\sqrt{\frac{1}{\left(\frac{1}{{\sigma^2}_1}+\frac{1}{{\sigma^2}_2}+\frac{1}{{\sigma^2}_3}+\frac{1}{{\sigma^2}_4}-\frac{2\rho_{12}}{\sigma_1\sigma_2}-\frac{2\rho_{13}}{\sigma_1\sigma_3}-\frac{2\rho_{14}}{\sigma_1\sigma_4}-\frac{2\rho_{23}}{\sigma_2\sigma_3}-\frac{2\rho_{24}}{\sigma_2\sigma_4}-\frac{2\rho_{34}}{\sigma_3\sigma_4}+\frac{1}{\tau^2}\right)}}}\right)^2+c'$$