COMPUTING WITH WORDS: FROM LINGUISTIC PREFERENCES TO DECISIONS

by,

FARAN AHMED

Submitted to the Graduate School of Engineering and Natural Sciences

in Partial Fulfillment of the Requirements for the Degree of

Ph.D

in

Industrial Engineering

Sabancı University

Spring 2019

COMPUTING WITH WORDS:
FROM LINGUISTIC PREFERENCES TO DECISIONS

APPROVED BY:

Doç. Dr. Kemal Kılıç ....
(Thesis Supervisor)

Dr. Öğr. Üyesi. Murat Kaya ....

Doç. Dr. Ayşe Kocabıyıkoğlu ....

Dr. Öğr. Üyesi. Ilker Köse ....

Doç. Dr. Gurdal Ertek ....

DATE OF APPROVAL: 07/01/2019

ABSTRACT

Computing with Words: From Linguistic Preferences to Decisions

Faran Ahmed
PhD in Industrial Engineering
Thesis Supervisor: Dr. Kemal Kılıç

Keywords: *Computing with Words, Linguistic Preferences, Pairwise Comparisons, Analytic Hierarchical Process*

Lexicons help to make qualitative assessments in various application areas such as Multi Criteria Decision Making (MCDM), intelligence analysis and human-machine teaming. In order to make quantitative analysis, these qualitative assessments based on the lexicons need to be quantified. During quantification, the linguistic descriptors involved in the lexicons that represent the judgments of the decision makers are mapped to a number. This is often achieved by using a fixed numeric scale. However, for a variety of reasons, such as the vagueness of the linguistic descriptors, the personal differences between the meanings associated to these linguistic descriptors, and the difference between the usage habits of the decision makers, it is not a realistic expectation to perform this mapping with a universal fixed numerical scale. Thus, many researchers frequently criticize this practice. In our study, we focused on the quantification of these linguistic descriptors. The performance of the different approaches used in quantification phase are comparatively assessed and various new proposals are made in order to improve the success of quantification process.

Although the quantification of qualitative assessments is a process that has been encountered in many different applications, in this study we have targeted the Analytic Hierarchical Process (AHP) framework, which is proposed by Thomas L. Saaty and widely used in MCDM. In AHP, the relative weights of the criteria and/or the utility of the alternatives for a criterion are determined from the qualitative assessments attained from the decision makers via pairwise comparisons. These qualitative assessments are quantified (often by Saatys 1-9 universal scale) in order to conduct further analysis. Thus, the quantification of the qualitative assessments, which can also contain various rational and/or irrational elements, is naturally a critical step for the success of the whole process.

In the scientific literature, various approaches are developed in order to improve the quantification process. Fuzzy AHP (FAHP), which integrates fuzzy set theory to the original AHP, is one of the most popular approach that is proposed for this purpose. Numerous FAHP algorithms were developed, which used fuzzy numbers as a scale to quantify the qualitative assessments.

However, there is no numerical or empirical study available that assesses the contribution of FAHP algorithms in MCDM. There is even no study, which evaluates the relative performances

of the FAHP algorithms and provide guideline to the researchers that frequently utilize these techniques as part of their analysis. Thus, in this study, firstly, the relative performances of the five popular FAHP algorithms, which are determined by number of citations they received in scientific literature, were measured by an experimental design study. In this context, four new FAHP algorithms were also developed and included in the experimental analysis. In the experimental analysis, three parameters, namely, the matrix size, the degree of inconsistency and the fuzzification parameter, were considered and the performance of the nine algorithms are assessed in various experimental conditions. This study revealed that the improved LLSM and the FICSM algorithm proposed in this study generally outperform the other algorithms significantly. To our surprise, the most popular algorithm in the literature, namely FEA, was the worst performing algorithm in the experimental analysis. On the other hand, the improved FEA significantly improved the performance of the original FEA.

In the second part of the study, the contribution of the FAHP algorithm in MCDM is discussed. Thomas L. Saaty himself criticized fuzzification of AHP arguing that judgments provided by experts are already fuzzy in nature and further fuzzifying them will add more inconsistency in the pairwise comparison matrices. Other researchers have made similar remarks mostly based on various theoretical arguments. However, these arguments are not supported by any numerical or empirical study. In this research, we addressed this gap as well and the contribution of FAHP to MCDM was investigated by means of numerical and empirical analysis. The FAHP algorithms, which outperformed the others in the first part of the study, were compared with the original AHP algorithms. The results revealed that the original AHP algorithms significantly outperformed the existing FAHP algorithms. The results of numerical and empirical analysis suggests that either the existing FAHP methods need to be improved or new ones should be developed in order to benefit the researchers working in MCDM.

In addition to the FAHP algorithms, another approach that aims to improve the quantification step is personalization of the numerical step instead of using a universal fixed scale. In the last part of the study we addressed this approach and investigate its performance. Two simple and intuitive heuristics are also developed as an alternative to the existing relatively complex mathematical programming based personalization approach since most of the researchers and practitioners utilizing MCDM techniques might not be familiar with optimization. Both the numerical analysis and the empirical studies demonstrated that the heuristic approaches outperformed the original AHP methods significantly.

ÖZET

Sözcüklerle Hesaplama: Sözel Değerlendirmelerden Kararlara

Faran Ahmed
Endüstri Mühendisliği Doktora Programı
Tez Danışmanı: Dr. Kemal Kılıç

Anahtar Sözcükler: *Sözcüklerle Hesaplama, Sözel Yargılar, Çift Yönlü Karşılatırmalar, Analitik Hiyerarşik Süreci*

Sözlükler, Çok Kriterli Karar Problemleri (ÇKKP), istihbarat analizi ve insan-makine takımları gibi çeşitli uygulama alanlarında nitel değerlendirmelerin yapılmasına yardımcı olur. Daha nicel analizlerin yapılması için sözcükler kullanılarak yapılan bu nitel değerlendirmelerin sayısallaştırılmaları gerekmektedir. Sayısallaştırma aşamasında karar vericilerin yargılarını ifade eden sözcükler bir sayıyla eşleştirilirler ve çoğu zaman bu sabit bir ölçek kullanılarak gerçekleştirilir. Ancak, sözcüklerin muğlaklığı, kişilerin sözcüklere yüklediği anlamların farklılığı ve kullanım alışkanlıkları gibi çeşitli nedenlerle, evrensel geçerliliği olan sabit bir sayısal ölçek ile bu eşleştirmenin yapılması gerçekçi bir beklenti değildir. Nitekim pek çok araştırmacı sıklıkla yapılan bu uygulamaya dönük eleştiriler dillendirilmektedir. Çalışmamızda sözcüklerin sayısallaştırılması sürecine odaklanılmış, bu süreçte kullanılan yaklaşımlar mukayeseli bir şekilde ele alınarak performansları değerlendirilmiş ve çeşitli iyileştirme önerileri geliştirilmiştir.

Nitel değerlendirmelerin sayısallaştırılması her ne kadar pek çok farklı uygulamada karşımıza çıkan bir süreç ise de, bu araştırmada kendimize ÇKKP kapsamında yaygın olarak kullanılan ve Thomas L. Saaty tarafından önerilen Analitik Hiyerarşik Süreç (AHP) çerçevesini hedef aldık. AHP kapsamında uzmanlardan gerek kıstasların (kriterlerin) göreceli ağırlıklarının belirlenebilmesi, gerekse her bir kıstas bağlamında alternatiflerin göreceli faydasının belirlenebilmesi amacıyla, nitel değerlendirmeleri ikili karşılaştırmalar şeklinde alınmaktadır. İçlerinde rasyonel veya irrasyonel ögeler de barındırabilen bu nitel değerlendirmelerin yapılacak analizlerde kullanılabilmesinin sağlayacak olan sayısallaştırılma aşaması doğal olarak sürecin bütünü kapsamında kritik önemdedir.

Bilimsel yazında bu aşamanın daha sağlıklı bir şekilde gerçekleştirilmesine yönelik çeşitli öneriler bulunmaktadır. Bulanık küme teorisinin, orijinal AHP'ye entegrasyonu sonucunda geliştirilmiş olan Bulanık AHP (FAHP), bu öneriler arasında yer alan ve oldukça sık karşımıza çıkan yaklaşımlardan birisidir. Nitel değerlendirmelerin sayısallaştırılması aşamasında kullanılan ölçeğin bulanık sayılardan oluşturulduğu, ardından bu bulanık sayılardan kıstasların ağırlığının veya her bir kıstas bağlamında alternatiflerin göreceli faydasının hesaplandığı çok sayıda FAHP algoritması geliştirilmiş ve bilimsel yazına sunulmuştur.

Sayısallaştırma aşamasında bulanık sayıların kullanıldığı ölçeğin, sürecin bütününe yaptığı katkının ölçülmesi bir yana, pek çok araştırmacının ÇKKP kapsamında kullandığı bu algoritmaların göreceli performansları dahi ölçülmemiştir. Bu araştırma kapsamında ilk olarak, bilimsel yazında yapılan atıflar gözetilerek belirlenen popüler beş FAHP algoritmasının göreceli performansları, bir deneysel tasarım çalışmasıyla ölçülmüştür. Bu kapsamda dört yeni FAHP algoritması da geliştirilmiş ve deneysel tasarım çalışmasına dâhil edilmiştir. Deneysel tasarım çalışmasında matris büyüklüğü, tutarsızlık derecesi ve bulanıklaştırma derecesi parametre olarak ele alınmış ve toplam dokuz algoritmanın farklı parametre seviyelerinde nasıl bir performans gösterdikleri ölçülmüştür. Bu çalışma bize iyileştirilmiş LLSM ve yeni geliştirilen FICSM algoritmalarının genellikle diğer algoritmalardan daha iyi çalıştığını göstermiştir. FEA gibi literatürde en popüler olan yaklaşımın ise en kötü sonuçları verdiği gene bu analizin sonucunda gözlemlenmiştir. İyileştirilmiş FEA ise orijinal FEA'in performansını önemli derecede arttırmıştır.

Araştırmanın ikinci kısmında ise bulanık sayılardan oluşan ölçeğin sürece yaptığı katkı ele alınmıştır. Thomas L. Saaty, bizzat kendisi bulanık sayılardan oluşacak olan ölçeğin faydasının olmayacağını zaten uzmanların niteliksel değerlendirmelerinin muğlak olduğu ve bulanık bir ölçeğin sürece daha da zarar vereceğini ifade etmiştir. Başka araştırmacılar da daha çok kuramsal argümanlarla benzer eleştirilerde bulunmuştur. Ama bu argümanlar ne deneysel tasarım çalışmalarıyla ne de gözlemsel çalışmalarla desteklenmemiştir. Bu çalışma kapsamında bu açığın giderilmesi hedeflenmiş ve hem deneysel tasarım çalışmasıyla hem de gözlemsel çalışmalarla FAHP'lerin sürece katkısının ölçülmüştür. Araştırmanın ilk kısmında belirlenen ve diğer FAHP algoritmalarına göre daha iyi performans sergileyen FAHP algoritmaları, orijinal AHP algoritmalarıyla kıyaslanmıştır. Sonuçlar, mevcut FAHP yöntemlerinin AHP yöntemlerinden daha kötü sonuçlar verdiklerini göstermiştir. Bu durum ise FAHP yöntemlerinin geliştirilmeleri gerektiğine, bu halleriyle kullanılmasının fayda değil zarar getirdiğine işaret etmektedir.

Nitel değerlendirmelerin sayısallaştırılmasında bulanık sayılardan oluşan bir ölçeği kullanan FAHP algoritmalarının yanı sıra evrensel bir ölçeğin yerine kişiselleştirilmiş bir ölçeğin kullanılmasına dayanan öneriler de bilimsel yazında yakın zamanda gündeme gelmiştir. Çalışmanın son kısmında bu yaklaşımın performansının ölçülmesi hedeflenmiştir. Bilimsel yazında bulunan, karmaşık ve pek çok araştırmacının üzerinde fazla bilgi sahibi olmadığı eniyileme yöntemlerine dayanarak önerilen kişiselleştirilme yaklaşımına alternatif olarak iki adet kolay sezgisel yaklaşım da bu kapsamda geliştirilmiştir. Gerek numerik deneysel tasarım çalışmasıyla, gerekse gözlemsel deneylerle geliştirilen sezgisel yaklaşımların bilimsel yazında bulunan orijinal AHP yöntemlerinden istatistiksel olarak ciddi derecede daha iyi bir performans gösterdiği gözlenmiştir.

This work is dedicated to

My Beloved Parents

Whose support, guidance and

encouragement have been the

source of inspiration throughout

completion of this project

Acknowledgments

Completion of this research would not have been possible without the valuable contribution, support and guidance, which I received throughout this project from various individuals. Therefore, I would like to take this opportunity to thank all those who made this research possible.

I would like to express my deepest gratitude to Dr. Kemal Kilic who offered his continuous advice and encouragement throughout the course of this dissertation. I thank him for his guidance and kind advisory services.

I also want to thank Dr. Murat Kaya, Dr. Ayşe Kocabiyikoğlu, Dr. Ilker Köse and Dr. Gurdal Ertek for accepting to be part of dissertation jury and their valuable feedback.

I gratefully acknowledge the funding received from Higher Education Commission of Pakistan to complete my MS and PhD degree.

Finally, I am thankful to my father Fahim Ahmed, my mother Tahira Fahim, my brother Sairam Ahmed and my wife Subha Khalid for their prayers, support and patience during this research work.

List of Abbreviations and Symbols

AHP	Analytic Hierarchy Process
CI	Consistency Index
CIV	Compatibility Index Value
CR	Consistency Ratio
FAHP	Fuzzy Analytic Hierarchy Process
FAM	Fuzzy Arithmetic Mean
FGM	Fuzzy Geometric Mean
FMRI	Functional Magnetic Resonance Imaging
FRSM	Fuzzy Row Sum Method
FNPCM	Fuzzy Numerical Pairwise Comparison Matrix
FLLSM	Fuzzy Logarithmic Least Square Method
FEA	Fuzzy Extent Analysis
FICSM	Fuzzy Inverse of Column Sum Method
INPCM	Individualized Numerical Pairwise Comparison Matrix
LLSM	Logarithmic Least Square Method
LPCM	Linguistic Pairwise Comparison Matrix
MCDM	Multiple Criteria Decision Making
NPCM	Numerical Pairwise Comparison Matrix
PNPCM	Polarized Numerical Pairwise Comparison Matrix
R.I	Random Index
TNPCM	Theoretical Numerical Pairwise Comparison Matrix
λ	Eigenvalue
μ	Membership Function
α	Fuzzification Parameter
β	Inconsistency parameter

Contents

List of Figures

List of Tables

Chapter 1

Background and Motivation

Lexicons that consists of linguistic qualifiers are used by experts to evaluate various situations (e.g., probabilities, significance etc.,) in business, academic, intelligence, medical, and political environments. These lexicons provide weakly ordered preferential relationship, which are used to evaluate alternatives and rank them accordingly. Elicitation of such linguistic qualifiers is of critical importance as it provides data which influences outcome in terms of rank, preferences scores or weights. Recent literature in neuroscience, especially in neuroeconomics, show that human preferences can be considered as a representation of the population of neuron activity [1]. There exists various psychological models of judgment and decision making which can be regarded as means to measure such brain activity and correspondingly interpret preferential relationships through real numbers.

First set of models are based on *Value First View* [2] [3] [4] [5] [6] and assume that preferences of objects expressed by decision makers through lexicons are associated with an internal numerical scale and brain compute values of available objects and chooses the one with higher values. Second set of models are based on *Comparison Based Theories* [7] [8] [9] [10], according to which, brain never computes anything in isolation, it rather computes how much it values one object compared to another object. Third set of models are based on *Comparison Based Theories without Internal Scale* [11] [12] [13] [14]. These models assume no values are computed during the comparisons and the process only involved ordinal comparisons. Majority of the models suggest that comparisons performed by the brain and resulting choices depend heavily on the internal numerical scale, if it exists, and the context of available options.

These psychological models of judgment require elicitation of human preferences in some form.

Review of relevant literature shows that, since uncertainty associated with human preferences is better communicated using vague verbal terms which are more intuitive and natural [15], people prefer to use linguistic qualifiers to express their opinions. Furthermore, precise numerical values are generally avoided because they may imply a sense of precision which a decision maker does not want due to uncertain nature of the whole decision problem [16]. For example, people mostly think and talk about uncertainty in terms of verbal phrases (*Extremely likely*, *not very likely* etc.,) and they are more skilled in using the rules of language as compared to the rules of probability [17]. These linguistic qualifiers are used across different domains and the ones used in multi-criteria decision making [18], intelligence [19], and medicine are tabulated in Table 1.1.

Table 1.1: Lexicons across different domains

Lexicons	**Numerical Equivalent**
Multi-Criteria Decision Making (MCDM)	
Equally Important	1
Moderately Important	3
Strongly Important	5
Very Strongly Important	7
Extremely Important	9
Words of Estimative Probability in Intelligence	
Almost Certain	93%
Probable	75%
Chances About Even	50%
Probably Not	30%
Almost Certainly Not	7%
Words of Estimative Probability in Buisness	
Likely	Expected to happen to more than 50 % of subjects
Frequent	Will probably happen to 10-50 % of subjects
Occasional	Will happen to 1-10 % of subjects
Rare	Will happen to less than 1 % of subjects

In this research we target lexicons that are used in Multi-Criteria Decision Making (MCDM) models which refer to decisions making in the presence of multiple, usually conflicting criteria. Analytic Hierarchy Process (AHP) proposed by Thomas L. Saaty [18] is one of the most popular methods in MCDM in which expert opinions are elicited in the form of pairwise comparisons. Making pairwise comparisons is the preferred way of eliciting such human preferences as it deals with binary evaluations and is an easier cognitive task when compared to evaluating all objects simultaneously [20].

In Pairwise comparisons, two objects are evaluated simultaneously and preference intensities are provided in the form of linguistic qualifiers which are then stored in Linguistic Pairwise

Comparison Matrices (LPCMs). Afterwards, the challenge is to compute with these LPCMs and derive priority vectors. As of now, linguistic qualifiers are transformed into numbers through a fixed numerical scale and the computations are carried out with these numbers. The most common approach in the literature is to use a fixed scale [21] [22] [23] [24] [25]. These scales are discussed in Section 2.1.

Two major criticisms associated with conventional fixed scales are; firstly, rather than quantifying linguistic variables with crisp numbers, fuzzy numbers might represent reality better, and secondly, a constant fixed scale for all individuals is not logical as words have different meaning for different people. In order to address first criticism, various Fuzzy AHP (FAHP) methods have been developed in which linguistic qualifiers are numerically represented through fuzzy numbers and priority vectors are derived from fuzzy pairwise comparison matrices. Logarithmic least squares method [26], modified logarithmic least squares method [27], geometric mean method [28] and fuzzy extent analysis [29] are the most commonly used FAHP methods in the literature.

In order to address the second criticism, a novel approach is proposed to generate personalized numerical scale, which utilizes a mathematical model to quantify linguistic qualifiers at an individual level [30]. The objective function of this model minimizes the inconsistency of the Numerical Pairwise Comparison Matrix (NPCM). However, efforts to reduce inconsistency in a pairwise comparison matrix can distort the meaning of linguistic qualifiers to an extent that they no longer represents decision makers preferences. Distinction should be made in the consistency of the preferences and the validity of the underlying decision process. Improving consistency of NPCM does not necessarily improves the validity of the results and thus consistency improving methods could be misleading [31].

In light of aforementioned arguments, we first present different numerical scales in Chapter 2. Details of different methods to derive a weight vector from numerical pairwise comparison matrices is discussed in Chapter 3 and 4. Various consistency measures are presented in Chapter 5. Based on the research methodology outlined in Chapter 6, a detailed performance analysis of nine FAHP methods is discussed in Chapter 7. Results from the evaluation of the value of fuzzifying human preferences is presented in Chapter 8. We propose two novel heuristics that will be used for a particular expert to quantify his/her linguistic qualifiers. These heuristics and their comparison with traditional methods are discussed in Chapter 9. We conclude this dissertation by providing concluding remarks and future research areas in Chapter 10.

Chapter 2

Numerical Scale

Eliciting human preferences in the form of pairwise comparisons is the preferred way of eliciting human preferences as it deals with binary evaluations and it is an easier cognitive task when compared to evaluating multiple objects simultaneously [20]. Eliciting preferences numerically may seem beneficial as they are more precise, permit communication to be less ambiguous and can be easily used in subsequent calculations [32]. However, review of relevant literature shows that people prefer to use verbal statements to express their opinions, as uncertainty associated with pairwise comparisons is better communicated using vague verbal terms which are more intuitive and natural [15]. Furthermore, precise numerical values are generally avoided because they may imply a sense of precision which a decision maker avoid due to uncertain nature of the decision problem [16]. People mostly think and talk about uncertainty in terms of words, i.e., linguistic labels and they are more skilled in using the rules of language as compared to the rules of probability [17]. Therefore, it is fair to conclude that the best way to elicit expert opinions is to use linguistic labels in the form of natural language verbal phrases.

Once decision makers and/or experts provide pairwise comparisons in the form of linguistic variables, this information is stored in Linguistic Pairwise Comparison Matrices (LPCMs). Next step in the process is to compute with these LPCMs in order to derive priority vectors for available criteria and alternatives. As of now, rather than computing with words, these linguistic variables are transformed into numbers and the computations are carried out with these numbers. The most common approach in the literature is to use a fixed numerical scale. There are two types of fixed scales, i.e., scale based on crisp numbers and scale based on fuzzy numbers. Recently, research is being conducted on personalized scale in which each linguistic label has a numerical value optimized at an individual level. In the following sections, different numerical scales to convert linguistic labels into numbers will be discussed in detail.

2.1 Fixed Scale

As stated above, one way of computing with linguistic labels is by transforming them into numbers using a fixed numerical scale. It maps linguistic labels to numbers and correspondingly LPCMs are converted into Numerical Pairwise Comparison Matrices (NPCMs). One such example is tabulated in Table 2.1.

Table 2.1: Linguistic variables transformed into numbers using Saaty scale of 1-9

Linguistic scale	**Linguistic Variable** S_i	**Numerical Value** $f(s_i)$
Equally Important	S8	1
Weakly More Important	S9	2
Moderately More Important	S10	3
Moderately Plus More Important	S11	4
Strongly More Important	S12	5
Strongly Plus More Important	S13	6
Demonstrated More Important	S14	7
Very Strongly More Important	S15	8
Extremely More Important	S16	9

Where S_i is a linguistic variable and it holds a value in the form of lexicons. For example, $S_8 =$ *Equally Important*, $S_9 =$ *Weakly More Important* etc.,. Linguistic variables S_0 to S_7 represent inverse of linguistic variables tabulated in Table 2.1. Function f transforms linguistic variables into numbers and inverse of function (f^{-1}) transform numbers into linguistic variables. For clarity, an example of LPCM and its transformation to NPCM is given as follows;

$$\begin{bmatrix} S_8 & S_9 & S_{15} \\ S_7 & S_8 & S_{11} \\ S_1 & S_5 & S_8 \end{bmatrix} = \begin{bmatrix} f(S_8) & f(S_9) & f(S_{15}) \\ f(S_7) & f(S_8) & f(S_{11}) \\ f(S_1) & f(S_5) & f(S_8) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 8 \\ 1/2 & 1 & 4 \\ 1/8 & 1/4 & 1 \end{bmatrix}$$

Besides the standard 1-9 scale, there are numerous other fixed scales proposed in the literature such as Linear Scale [18], Power Scale and scale based on Root Square [21], Geometric Scale [22], Inverse Linear [33], Asymptotic Scale [23], Balanced Scale [24] and Logarithmic Scale [34] etc.,. Details of these scales are tabulated in Table 2.2 and comprehensive review of these scales in terms of sensitivity to consistency and variance of allocation of priority values is provided in [35].

Table 2.2: Most common fixed scales used in AHP

Scale Type	Mathematical Description	Parameters
Linear	$s = x$	$x = 1, 2, ..., 9$
Power	$s = x^2$	$x = 1, 2, ..., 9$
Root Square	$s = \sqrt{x}$	$x = 1, 2, ..., 9$
Geometric	$s = 2^{x-1}$	$x = 1, 2, ..., 9$
Inverse Linear	$s = \frac{9}{10-x}$	$x = 1, 2, ..., 9$
Asymptotical	$s = tanh^{-1}\left(\frac{\sqrt{3}(x-1)}{14}\right)$	$x = 1, 2, ..., 9$
Balanced	$s = \frac{x}{1-x}$	$x = 0.5, 0.55, 0.6, ..., 9$
Logarithmic	$s = log_2(x + 1)$	$x = 1, 2, ..., 9$

It has been argued in the literature that human preferences are vague in nature and representing them with a constant fixed scale cannot incorporate the inherent uncertainty associated in human observations. Furthermore, discreteness and finiteness of these scales is one of the primary sources of inconsistency in NPCMs [36]. Different approaches have been proposed in the literature that attempts to address these issues. One such approach is to represent human preferences through fuzzy numbers, which is explained in Section 2.2 and Chapter 4. Another approach is to use a personalized scale that construct numerical scales for each individual separately. This approach will be discussed in Section 2.3.

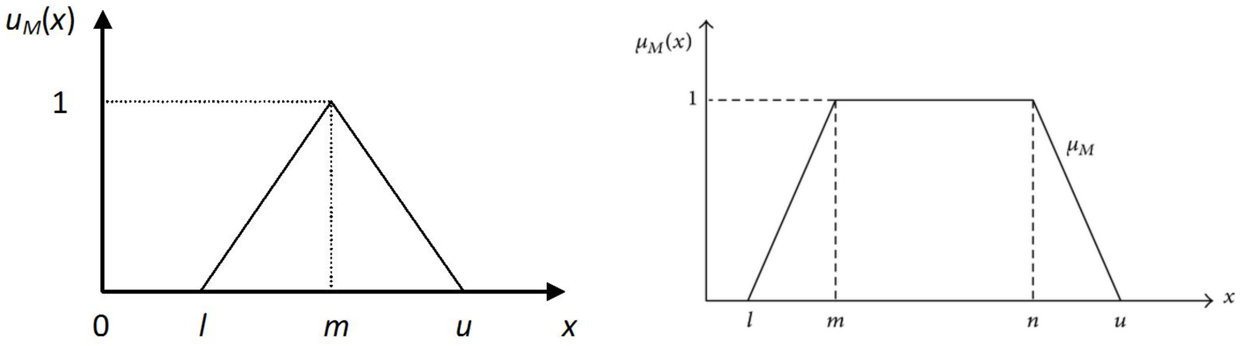
2.2 Scale Based on Fuzzy Numbers

The main motivation behind using fuzzy set theory for mapping linguistic labels is based on the argument that human judgment and preferences cannot be accurately represented by crisp numbers due to the inherent uncertainty in human perception. Disregarding this fuzziness of the human behavior in the decision making process may lead to wrong decisions [37]. In order to address this issue of vagueness and uncertainty, fuzzy set theory introduced by Zadeh [38] has been extensively incorporated into the original AHP in which the weighing scale is composed of fuzzy numbers. In this section, we present a brief overview of fuzzy numbers, fuzzy arithmetics and fixed fuzzy scale and later in Chapter 4, various techniques associated with fuzzy scales will be presented.

Instead of a single value, a fuzzy number represents a set of possible values each having its own membership function varying between zero and one. A triangular fuzzy number is represented by *[lower value, mean value, upper value]* or $[l\ m\ u]$ and a trapezoidal number is represented by $[l\ m\ n\ u]$ with membership functions μ_M given by;

$$\mu_M(x) = \begin{cases} \frac{x}{m-l} - \frac{l}{m-l}, & x \in [l\ m] \\ \frac{x}{m-u} - \frac{u}{m-u}, & x \in [m\ u] \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

Note that the membership function defined in Equation 2.1 is for triangular fuzzy numbers. For trapezoidal numbers, membership function in the interval $[m\ n]$ is equal to one. The same is graphically illustrated in Figure 2.1.



(a) Triangular Fuzzy Number

(b) Trapezoidal Fuzzy Number

Figure 2.1: Membership function of fuzzy numbers

An example of fixed scale based on triangular fuzzy number (TFN) is illustrated in Table 2.3. Let $(l_1\ m_1\ u_1)$ and $(l_2\ m_2\ u_2)$ be two triangular fuzzy numbers and $(l_1\ m_1\ n_1\ u_1)$ be a trapezoidal fuzzy number, than the basic fuzzy arithmetic operations are tabulated in Table 2.4.

Table 2.3: An example of fixed scale based on triangular fuzzy number

Linguistic scale	Scale Based on TFN
Just Equal	(1, 1, 1)
Equally Important,	(1/2, 1, 3/2)
Weakly More Important,	(1, 3/2, 2)
Strongly More Important	(3/2, 2, 5/2)
Very Strongly More Important	(2, 5/2, 3)
Absolutely More Important	(5/2,3, 7/2)

Table 2.4: Fuzzy arithmetics

Operation	Result
Addition	$(l_1\ \ m_1\ \ u_1) \oplus (l_2\ \ m_2\ \ u_2) = (l_1+l_2\ \ m_1+m_2\ \ u_1+u_2)$
Multiplication	$(l_1\ \ m_1\ \ u_1) \odot (l_2\ \ m_2\ \ u_2) = (l_1.l_2\ \ m_1.m_2\ \ u_1.u_2)$
Scalar Multiplication	$\lambda \odot (l\ \ m\ \ u) = (\lambda .l\ \ \lambda .m\ \ \lambda .u)$
Inverse (Triangular Fuzzy Number)	$(l\ \ m\ \ u)^{-1} = (1/u\ \ 1/m\ \ 1/l)$
Inverse (Trapezoidal Fuzzy Number)	$(l\ \ m\ \ n\ \ u)^{-1} = (1/m\ \ 1/l\ \ 1/u\ \ 1/n)$

2.3 Personalized Scale

Transforming linguistic labels into numbers using any of the fixed scales discussed in the previous sections is based on the assumption that these linguistic variables have the same meaning for all individuals which is not the case in reality, as words have different meaning for different people [16]. A novel approach [30] is proposed to transform LPCMs into NPCMs which utilizes a mathematical model to generate personalized scale for each decision maker separately. The proposed mathematical model utilizes information provided by the decision maker through LPCMs and transitivity rules of pairwise comparisons to generate personalized scales for each individual. A brief overview of this mathematical formulation is provided as follows.

Suppose there are three criteria $\{C_1, C_2, C_3\}$ and pairwise comparisons provided by the decision maker between criterion C_1 and C_2 is s_r, between C_2 and C_3 is s_t and C_1 and C_3 is e_{rt}. Based on these pairwise comparisons, corresponding LPCM can be formulated as follows;

$$\begin{bmatrix} Null & S_r & e_{rt} \\ Null & Null & S_t \\ Null & Null & Null \end{bmatrix}$$

Let $\{\frac{1}{f_j}, f_1, f_j\}$ for $j = 2, 3, ..., 9$ be the numerical scale used to transform above LPCM into NPCM. As stated earlier, function f transforms linguistic variables into numbers and inverse of function i.e., f^{-1} transform numbers into linguistic variables. Then, based on a selected scale, the corresponding NPCM can be represented as follows;

$$\begin{bmatrix} Null & f(S_r) & f(e_{rt}) \\ Null & Null & f(S_t) \\ Null & Null & Null \end{bmatrix}$$

e_{rt} is the preference provided by the decision maker, however, the same can be estimated through the transitivity conditions for a perfectly consistent matrix i.e., $f(S_r) \times f(S_t) = f(e_{rt})$. Therefore, $\widehat{e}_{rt}$ can be estimated using Equation 2.2;

$$\widehat{e}_{rt} = f^{-1}(f(S_r) \times f(S_t)) \tag{2.2}$$

Selecting different numerical scales will yield different values for estimated $\widehat{e}_{rt}$. For example, if given preference intensities are S_9 and S_{10} then using Saaty scale of 1-9, $\widehat{e}_{9,10}$ can be estimated as follows;

$$\widehat{e}_{9,10} = f^{-1}(f(S_9) \times f(S_{10})) = f^{-1}(2 \times 3) = S_{13}$$

The main objective of the proposed model is to construct such a scale that minimizes the deviation between e_{rt} provided by the decision maker and $\widehat{e}_{rt}$ estimated from a particular numerical scale. Mathematical formulation of this model is given as follows;

$$\begin{aligned} \underset{f_2,\dots,f_g}{\text{minimize}} \quad & \sum_{r,t=0}^{16} d_(e_{rt}, \widehat{e}_{rt}) \\ \text{subject to} \quad & L_i \le f_i \le U_i, \quad i = 1, 2, \dots, 9, \\ & f_i < f_{i+1}, \qquad i = 1, 2, \dots, 8 \end{aligned} \tag{2.3}$$

Where e_{rt} represents linguistic information provided by the decision maker and $\widehat{e}_{rt}$ represents estimated linguistic information based on transitivity rules. In this research, this mathematical model to generate personalized scale for each individual will be investigated through a numerical and empirical study.

Chapter 3

Deriving a Weight Vector from Crisp NPCMs

After constructing NPCM using one of the scales discussed in Chapter 2, the next step is to identify the priority vector from NPCM. If crisp numbers are preferred in NPCM, then conventional AHP methods will be utilized to derive a priority vector from NPCM. On the other hand, if fuzzy numbers are preferred, FAHP methods will be used to derive the required priority vector. Over the years various priority derivation techniques have been proposed for both original AHP and FAHP methods. In this Chapter, we will discuss various priority derivation techniques in original AHP methods.

3.1 Eigenvector

Saaty [18] proposed that the principal eigenvector of the NPCM accurately estimates the desired priority vector. Provided we have a fully consistent comparison matrix and multiply it with the column priority vector (which we are trying to identify) we end up with following:

$$\begin{pmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = n \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \tag{3.1}$$

Therefore, given a comparison matrix A, we can solve for the priority vector $A \times w = n \times w$. In standard linear algebra, n in this equality is referred as eigenvalue and w is referred to as the corresponding eigenvector. That is to say, the eigenvector of a fully consistent PCM is the required priority vector. Note that as a general rule, sum of the eigenvalues of a $n \times n$ matrix A is equal to the trace i.e., sum of the diagonal elements of A. Due to the special structure of the fully consistent comparison matrix (i.e., the transitivity rule holds and as a result the rank of such a matrix is 1), it has only one eigenvalue and its value is n (the sum of the diagonal elements, $\sum_{i=1}^{n} 1 = n$).

In reality, we do not encounter a fully consistent comparison matrix assessed from the decision maker(s). Therefore, the comparison matrix yields multiple number of eigenvalues with values that are not equal to n. Saaty proposes to use the maximum eigenvalue among the set of the eigenvalues that would be obtained from a inconsistent comparison matrix, which would be closer to the theoretical value of n obtained from a fully consistent comparison matrix. Mathematical formulation for estimating maximum eigenvalues is given by following Equation.

$$A \times w = \lambda_{max} \times w$$

where $\lambda_{max} \approx n$. As explained earlier, in case of a perfectly consistent matrix $\lambda_{max} = n$. Once the eigenvector corresponding to the maximum eigenvalue is calculated, it is then normalized to estimate the final priority vector.

3.2 Logarithmic Least Squares Method

Let's assume that w_i and w_j are weights to be estimated while a_{ij} is the comparison ratio provided by the expert while comparing criterion i with criterion j. Due to the inherent inconsistency in human judgments, comparison ratio a_{ij} will differ from the corresponding set of weights. Therefore, the goal is to estimate such a combination of weights that minimizes the total deviation between comparison ratios provided by the expert and the ratio of the corresponding weights which can be achieved by minimizing following equation.

$$min \sum_{i<j} (\ln a_{ij} - (\ln w_i/w_j))^2 \tag{3.2}$$

Crawford and Williams [39] shows that solution to above problem is unique and can be found by taking geometric means of the rows of pairwise comparison matrix A.

3.3 Arithmetic and Geometric Mean Heuristic

Arithmetic and geometric mean approaches are among the most popular techniques used in original AHP. These two techniques originate from the properties of a fully consistent comparison matrix. Suppose there are (n) criteria and the aim is to extract the weight vector, i.e., $w = w_1, w_2, \cdots, w_n$ whereas w_i refers to the weight of the ith criteria. Recall that a fully consistent comparison matrix is as follows:

$$W' = \begin{pmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{pmatrix} \tag{3.3}$$

In the first step, elements of each columns are added, which results in the following;

$$\left(\frac{w_1 + w_2, \cdots, w_n}{w_1}, \frac{w_1 + w_2, \cdots, w_n}{w_2}, \cdots, \frac{w_1 + w_2, \cdots, w_n}{w_n} \right) \tag{3.4}$$

Note that AHP assumes additive utility, that is to say, the overall utility is weighted sum of individual utilities i.e., $\sum w_i = 1$. Therefore, the column sums provided by Equation 3.4 are equivalent to

$$\left(\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n}\right) \tag{3.5}$$

Next, if each element of the comparison matrix is divided by its corresponding column sum, we end up with the following matrix which is referred to as W^N.

$$W^N = \begin{pmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \ddots & \vdots \\ w_n & w_n & \cdots & w_n \end{pmatrix} \tag{3.6}$$

Hence, for a fully consistent matrix, if the above described normalization process is applied, the resulting matrix W^N will be composed of column vectors which are same with each other and any one of them can be considered as the weight vector that we to be determined i.e., $(w_1, w_2, \cdots, w_n)$. However, since in practice the comparison matrix obtained from the decision makers are rarely consistent, the resulting normalized matrix would be composed of column vectors that are different from each other. Since each column is a candidate for the weight vector and the source of the inconsistency cannot be detected, the reasonable approach is to average the columns of the normalized matrix W^N. The average can either be obtained by arithmetic means or geometric means. Equations 3.7 and 3.8 represent these two approaches, where w_i^J denotes the candidate weight associated with the i^{th} criteria based on the j^{th} column of W^N.

$$A.M = \frac{\sum_{J=1}^{n} w_i^J}{n} \quad for\ i = 1, 2, \cdots, n \tag{3.7}$$

$$G.M = \left[\prod_{J=1}^{n} w_i^J\right]^{1/n} \quad for\ i = 1, 2, \cdots, n \tag{3.8}$$

3.4 Row Sum Heuristic

Starting with the perfectly consistent comparison matrix given by Equation 3.3, we first take sum of all elements in i^{th} row and assign it to $R.S_i$. The sum of each row is as follows;

$$R.S_1 = w_1\left(\frac{1}{w_1} + \frac{1}{w_2} + \cdots + \frac{1}{w_n}\right)$$

$$R.S_2 = w_2\left(\frac{1}{w_1} + \frac{1}{w_2} + \cdots + \frac{1}{w_n}\right)$$
$$\vdots$$
$$R.S_n = w_n\left(\frac{1}{w_1} + \frac{1}{w_2} + \cdots + \frac{1}{w_n}\right)$$

The sum of all $R.S_i$ is given as follows;

$$\sum_{i=1}^{n} R.S_i = (w_1 + w_2 + \cdots + w_n).\left(\frac{1}{w_1} + \frac{1}{w_2} + \cdots + \frac{1}{w_n}\right)$$

where $(w_1 + w_2 + \cdots + w_n) = 1$ due to additive utility. Priority vector can be calculated by normalizing each $R.S_i$ by dividing it by $\sum_{i=1}^{n} R.S_i$. The priority vector is given as follows;

$$\frac{R.S_1}{\sum_{i=1}^{n} R.S_i} = w_1, \frac{R.S_2}{\sum_{i=1}^{n} R.S_i} = w_2, \cdots \frac{R.S_n}{\sum_{i=1}^{n} R.S_i} = w_n$$

3.5 Inverse of Column Sum

Another heuristic that is developed and included in this research is referred to as Inverse of Column Sums. This intuitive heuristic requires very few arithmetic operations. For the fully consistent comparison matrix, column sum of each column is calculated by Equation 2. Since (due to additive utility assumption) $w_1 + w_2 + \cdots + w_n = 1$, the column sums are equivalent to

$$\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n} \tag{3.9}$$

Thus, the inverse of column sums will yield the weight vector $w_1, w_2, \cdots, w_n$ for fully consistent comparison matrix.

3.6 Comparison of the Original AHP Methods

Proposing algorithms to derive accurate priority vector from inconsistent NPCMs have always been a fertile and popular research area for researchers working in this domain. Some of the techniques in original AHP; in addition to those stated previously are; Additive Normalization Method [40], Modified Eigenvector Method [41], Weighted Least-Squares Method [42], Logarithmic Goal Programming [43] and many others which are not discussed in detail due to space limitation. Although there exists numerous techniques to derive priority vector, Saaty argues that when comparison matrices are inconsistent, their transitivity effects the final outcome, and must be taken into account while deriving priority vector. As principal eigenvector (i.e., eigenvector associated with largest eigenvalue) captures transitivity uniquely, therefore, it is the proper way of obtaining accurate priority vector [44, 45]. Furthermore, inconsistency is an inevitable phenomenon in any NPCM elicited from decision maker and slight variations caused in inconsistent NPCMs are accordingly represented in slight variations in eigenvector and eigenvalues [31].

On the other hand, comparative analysis of 18 different techniques shows that distance-based estimating methods, such as least squares and various extensions of least squares are preferred as they minimize distance between matrix $[w_i/w_j]$ to the original comparison matrix provided by the decision maker [46]. Furthermore, it has been shown that the priority vector calculated through eigenvector approach can violate conditions of order preservation [47] and thus caution should be observed while using this technique to derive priority vector.

Golany and Kress [48] provide an analysis amongst six methods in which they use minimum violation, total deviation, conformity and robustness as criteria for performance analysis and concluded that Modified Eigenvalue (MEV) is the most ineffective method, while among the remaining five algorithms, each have their own weaknesses and advantages. Another comparative analysis performed by Ishizaka and Lusti [49] used Monte Carlo simulations to compare and evaluate four priorities deviation techniques which includes right eigenvalue method, left eigenvalue method, geometric mean and the mean of normalized values and conclude that number of contradictions increases with increase in the inconsistency as well as the size of the matrix. Some other similar studies are also available in the literature [50] [51] [52] [53] [54].

Chapter 4

Deriving a Weight Vector from Fuzzy NPCMs

In this chapter, we will discuss various priority derivation techniques, when the scale used to transform LPCM into NPCM is composed of fuzzy numbers. Resulting NPCMs are considered as fuzzy NPCMs (FNPCM) and we refer to FAHP when such a transformation is used to convert linguistic variables into numbers. Main motivation behind incorporating fuzzy set theory into original AHP is based on the argument that human judgments and preferences cannot be accurately represented by crisp numbers due to the inherent uncertainty in human perception. Therefore, in order to address this issue of vagueness and uncertainty, and to accurately transform human judgments into ratio scales, fuzzy set theory introduced by Zadeh [38] has been extensively incorporated into the original AHP in which the weighing scale is composed of fuzzy numbers.

In FAHP, weights are calculated from fuzzy comparison matrices which are later used to rank the available alternatives together with the scores attained by the alternatives for each criterion. Therefore, determination of the weights from comparison matrices is one of the key steps of the process. In the conventional AHP these weights are shown to be the eigenvectors of the comparison matrix for a fully consistent decision maker [45]. However in case of FAHP, calculating weights from fuzzy comparison matrices is not straightforward due to complexities associated with the arithmetics of fuzzy numbers. Therefore, over the past couple of decades, various FAHP methods have been proposed in the literature with an aim to accurately extract weights from fuzzy comparison matrices.

Golany and Kress [48] provide a performance comparison of most commonly used six methods in original AHP, in which they used minimum violation, total deviation, conformity and robustness as criteria for performance analysis. They concluded that Modified Eigenvalue (MEV) [41] is the least effective method, while among the remaining five algorithms, each have their own weaknesses and advantages. Another comparative analysis is performed by Ishizaka and Lusti [49] in which they used Monte Carlo simulations to compare and evaluate four techniques to derive priority vector in original AHP including right eigenvalue method, left eigenvalue method, geometric mean and the mean of normalized values and conclude that number of contradictions increases with increase in the inconsistency as well as the size of the matrix. Some other similar studies that compare crisp AHP approaches in various aspects are also available in the literature [50] [51] [52] [53] [54].

Buyukozkan et al. [55] provide a review of FAHP methods and list the characteristics and advantages and disadvantages for those methods which are structurally different. However, a *performance analysis* of FAHP methods similar to the ones that are available for the original AHP is not conducted so far. More and more papers are being published which apply FAHP as part of the solution process; however, the choice of the FAHP technique used in the analysis seems to be arbitrary due to this gap in the literature. Therefore, in this study we attempt to carry out a detailed performance analysis of nine different FAHP methods in terms of accuracy of weights calculated from fuzzy comparison matrices. Such an analysis would address the gap in the literature and guide both the researchers and practitioners while choosing the most appropriate FAHP method in their analysis.

As stated earlier, there are numerous FAHP methods proposed in the literature by various authors to effectively derive priority vector from FNPCM. The seminal article integrating fuzzy set theory and AHP was written by Van Laarhoven and Pedrycz [26] in which they utilized triangular fuzzy numbers and implemented logarithmic least squares method (LLSM) to derive priority vector from FNPCMs. Other popular FAHP techniques includes Modified Logarithmic Least Squares Method [27], Geometric Mean [28] and Fuzzy Extent Analysis [29]. Figure 4.1 illustrates the yearly citation history of these algorithms in Google scholar between 2000 and 2017. Review of this literature shows that some of these techniques (particularly Chang, Laarhoven and Buckley) are still being frequently referred to (and in some cases implemented as well) by the researchers. Brief overview of these techniques are presented in the following sections;

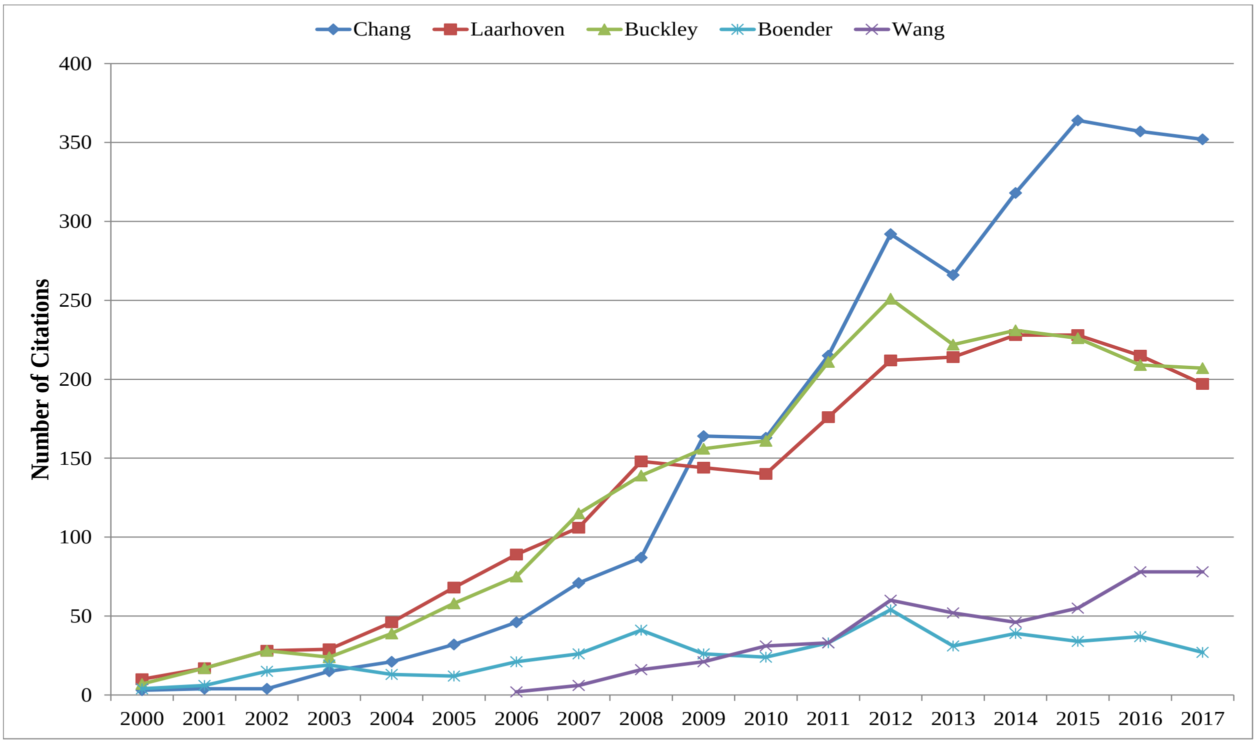


Figure 4.1: Number of citations received by the five popular algorithms in Google scholar between years 2000 and 2017

4.1 Fuzzy Logarithmic Least Squares Method

Van Laarhoveen and Pedrycz [26] suggested one of the first models in the domain of Fuzzy AHP, which utilizes fuzzy logarithmic least squares method (LLSM) and formulated an unconstrained optimization model to obtain triangular fuzzy weights. Note that Equation 3.2 presented in Section 3.2 is valid when comparison ratios are provided by a single expert and can be rewritten for multiple experts as follows.

$$min \sum_{i<j} \sum_{k=1}^{\delta_{ij}} (\ln a_{ijk} - (\ln w_i/w_j))^2 \tag{4.1}$$

where, δ_{ij} is the number of comparison ratios assessed from different experts available for a certain criteria. Equation 4.1 is simplified by replacing $y_{ijk} = \ln a_{ij}$, $x_i = \ln w_i$ and $x_j = \ln w_j$;

$$min \sum_{i<j} \sum_{k=1}^{\delta_{ij}} (y_{ijk} - x_i + x_j)^2 \tag{4.2}$$

To minimize Equation 4.2, we take partial derivatives with respect to x_i and equate them to zero. Following is the resultant set of equations.

$$x_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} x_j = \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=1}^{n} y_{ijk} \tag{4.3}$$

The above system is composed of set of equations which can be simultaneously solved to calculate all x_i's. Afterwards, to convert the system into its original form, exponential of the solution are taken and then normalized to estimate final weights. However, system of Equations presented in (4.3) is applicable only when the given comparison ratios are in the form of crisp numbers. It can be transformed for triangular fuzzy weights while following the rules for fuzzy arithmetic operations presented earlier in Table 2.4. This transformation is given as follows;

$$l_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} u_j = \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=1}^{n} l_{ijk} \tag{4.4}$$

$$m_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} m_j = \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=1}^{n} m_{ijk} \tag{4.5}$$

$$u_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \delta_{ij} l_j = \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=1}^{n} u_{ijk} \tag{4.6}$$

where $l_i = \ln w_{il}$, $m_i = \ln w_{im}$ and $u_i = \ln w_{iu}$. Same procedure is followed to convert the system into its original form by taking exponential of the solutions and then normalizing to estimate final fuzzy weights;

$$\tilde{w_i} = \left(\frac{exp(l_i)}{\sum_{i=1}^{n} exp(u_i)}, \frac{exp(m_i)}{\sum_{i=1}^{n} exp(m_i)}, \frac{exp(u_i)}{\sum_{i=1}^{n} exp(l_i)} \right) \tag{4.7}$$

4.2 Modified Fuzzy Logarithmic Least Squares Method

Subsequent research on this model identifies various irregularities and appropriate modifications are proposed [27, 56]. In the original LLSM Model, normalization process eliminates the optimality in the sense that the normalized solution violates the first order optimality conditions and thus normalized weights do not minimize the objective function. A modified version of the normalization procedure is proposed by Boender et al. [27] as follows;

$$\tilde{w}_i = \left(\frac{exp(l_i)}{\sqrt{\sum_{i=1}^{n} exp(l_i). \sum_{i=1}^{n} exp(u_i)}}, \frac{exp(m_i)}{\sum_{i=1}^{n} exp(m_i)} \frac{exp(u_i)}{\sqrt{\sum_{i=1}^{n} exp(l_i). \sum_{i=1}^{n} exp(u_i)}} \right) \tag{4.8}$$

4.3 Constrained Nonlinear Optimization Model

Wang et.al [56] criticize various aspects of the FLLSM Model and propose a constrained nonlinear optimization model to address those criticisms. His criticism of the original LLSM model is summarized as follows.

Incorrect Normalization: Fuzzy weights calculated after normalization procedure must satisfy the following conditions [57].

$$\begin{aligned} \sum_{i=1}^{n} w_i^U - \max_j (w_j^U - w_j^L) \geq 1 \\ \sum_{i=1}^{n} w_i^M = 1 \\ \sum_{i=1}^{n} w_i^L - \max_j (w_j^U - w_j^L) \leq 1 \end{aligned} \tag{4.9}$$

Although the normalization procedure modified by [27] provides optimal weights, [56] shows a counter example in which normalized fuzzy weights violate the conditions presented in Equation 4.9.

Incorrectness of Triangular Fuzzy Weights: The solution to the original system of equations can be represented as $(l_i + p_1, m_i + p_2, u_i + p_1)$. It was stated by [26] that arbitrary parameters p_1 and p_2 can be always chosen in a way that will ensure that the following condition is satisfied;

$$l_i + p_1 \leq m_i + p_2 \leq u_i + p_1, \quad for\ i = i, ..., n$$

After taking exponential and normalizing, fuzzy weights are again in the correct order. However, this claim was found false and a counter example was provided by [56] in which the normalized solution violated the given condition of a triangular fuzzy number.

Uncertainty of fuzzy weights for incomplete comparison matrices: In case of a comparison matrix in which some of the values/ratios are missing, the system of equations formed may contain free variables. Therefore, different configurations of free variables must be formed with each configuration leading to different weights. Such a situation is observed in the numerical example cited by [27]; however no justification is provided for choosing a specific configuration. This uncertainty in estimating fuzzy weights exists in all incomplete fuzzy PCMs and thus should be addressed appropriately.

In light of the above, [56] suggests a constrained nonlinear optimization model as follows:

$$minJ = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \sum_{k=1}^{\delta_{ij}} (\ln w_i^L - \ln w_j^U - \ln a_{ijk}^L)^2 + (\ln w_i^M - \ln w_j^M - \ln a_{ijk}^M)^2 + (\ln w_i^U - \ln w_j^L - \ln a_{ijk}^U)^2$$

$$\text{Subject to} \quad \begin{cases} w_i^L + \sum_{j=1, j\neq i}^{n} w_j^U \geq 1 \\ w_i^U + \sum_{j=1, j\neq i}^{n} w_j^L \leq 1 \\ \sum_{i=1}^{n} w_i^M = 1 \\ \sum_{i=1}^{n} (w_i^L + w_i^U) = 2 \\ w_i^U \geq w_i^M \geq w_i^L \end{cases} \tag{4.10}$$

A solution to this mathematical model is normalized fuzzy weights for both complete and incomplete comparison matrices. The first three constraints in (4.10) satisfy the normalization conditions of fuzzy numbers, the fourth constraint ensures a unique solution and the last constraint ensures that the condition $l < m < u$ is always satisfied.

4.4 Fuzzy Extent Analysis

Provided that $X = \{x_1, x_2, \cdots, x_n\}$ represents an object set and $G = \{g_1, g_2, \cdots, g_n\}$ represents a goal set, then as per the extent analysis method [29], for each object, extent analysis for each goal g_i is performed. Applying this theory in fuzzy comparison matrix, we can calculate value of fuzzy synthetic extent with respect to the i^{th} object as follows;

$$S_i = \sum_{j=1}^{m} M_{g_i}^{j} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^{j} \right]^{-1} \tag{4.11}$$

Where

$$\sum_{j=1}^{m} M_{g_i}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j \right) \tag{4.12}$$

Later in the decision making process (i.e., choosing the best alternative) we need to determine a crisp weight from these fuzzy triangular weights. A naive approach would be just using the means (i.e., mean of each fuzzy weight obtained from Equation 4.4). However, as opposed to the straightforward ordering of crisp numbers, the orderings of the fuzzy numbers are not that simple and one should be more careful. Chang [29] suggest utilizing the concept of comparison of fuzzy numbers in order to determine crisp weights from the fuzzy weights. In their approach, for each fuzzy weight, a pair wise comparison with the other fuzzy weights are conducted, and the degree of possibility of being greater than these fuzzy weights are obtained. The minimum of these possibilities are used as the overall score for each criterion i. Finally these scores are normalized (i.e., so that they sum up to 1), and the corresponding normalized scores are used as the weights of the criteria. That is to say by applying the comparison of the fuzzy numbers, the degree of possibility is obtained for each pair wise comparison as follows:

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2}(d) = \begin{cases} 1, & \text{if } m_2 \geq m_1 \\ 0, & \text{if } l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & \text{otherwise.} \end{cases}$$

The same is illustrated in the Figure 4.2.

Degree of possibility for a convex fuzzy number to be greater than k convex fuzzy numbers is given by;

$$V(M \geq M_1, M_2, \cdots, M_k) = V[(M \geq M_1)\text{and}(M \geq M_2), \cdots, (M \geq M_k)] \tag{4.13}$$

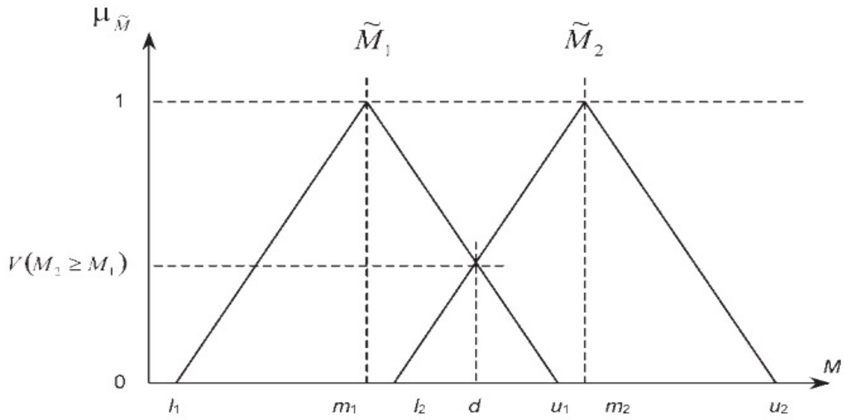


Figure 4.2: Degree of possibility

$$= \min V(M \geq M_i), \quad i = 1, 2, \cdots, k \tag{4.14}$$

Assuming that $w'_i = \min V(M_i \geq M_k)$ then weight vector is given by

$$W' = w'_1, w'_2, \cdots, w'_n \tag{4.15}$$

Normalizing the above weights gives us the final priority vector $w_1, w_2, \cdots, w_n$. Wang et.al [56] criticized fuzzy extent analysis technique and through an example showed that this method cannot estimate true weights from fuzzy comparison matrix. Their main criticism revolves around the fact that this method may assign a zero as criterion weight which disturbs the whole decision making hierarchy. The basis of extent analysis theory is that it provides a degree to which one fuzzy number is greater than another fuzzy number, and this degree of greatness is considered as criterion weights. Therefore, if two fuzzy numbers do not intersect then the degree of greatness of one fuzzy number to the other is 100 percent and therefore it will assign 1 as weight to that criterion while the other criteria will be assigned as zero weight. However, this still remains as one of the most popular FAHP technique used by the practitioners in their decision making problems.

4.5 Buckley Geometric Mean Method

Geometric mean method was proposed by Buckley [28] in which, instead of triangular fuzzy numbers, trapezoidal numbers were used to represent linguistic variables. Trapezoidal numbers are defined by $(l/m, n/u)$ where $0 < l \leq m \leq n \leq u$ and their membership function is illustrated in Figure 2.1. Expert judgment is recorded in a comparison matrix by fuzzy ratio $a_{ij} = (l_{ij}/m_{ij}, n_{ij}/u_{ij})$ whereas $l, m, n, u \in \{1, 2, \cdots, 9\}$. Following calculations are required in order to estimate final weight vector.

$$l = \sum_{i=1}^{n} l_i \quad \text{where as} \quad l_i = \left[\prod_{j=1}^{n}\right]^{1/n} \tag{4.16}$$

$$m = \sum_{i=1}^{n} m_i \quad \text{where as} \quad m_i = \left[\prod_{j=1}^{n}\right]^{1/n} \tag{4.17}$$

$$n = \sum_{i=1}^{n} n_i \quad \text{where as} \quad n_i = \left[\prod_{j=1}^{n}\right]^{1/n} \tag{4.18}$$

$$u = \sum_{i=1}^{n} u_i \quad \text{where as} \quad u_i = \left[\prod_{j=1}^{n}\right]^{1/n} \tag{4.19}$$

The final priority vector is given by $\left(\frac{l_i}{u}, \frac{m_i}{n}, \frac{n_i}{m}, \frac{u_i}{l}\right)$ and the corresponding membership function of the resulting trapezoidal fuzzy number is given by;

$$f_i(y) = \left[\prod_{j=1}^{n}((m_{ij} - l_{ij})y + l_{ij})\right]^{1/n} \tag{4.20}$$

$$g_i(y) = \left[\prod_{j=1}^{n}((n_{ij} - u_{ij})y + u_{ij})\right]^{1/n} \tag{4.21}$$

For complete details on this methodology, kindly refer to [28].

4.6 Fuzzification of Original AHP Heuristics

AHP heuristics explained in section 3.3, 3.4 and 3.5 are extended in the fuzzy case by replacing conventional arithmetic operations with fuzzy arithmetic operations. Let $(l_1\ m_1\ u_1)$ and $(l_2\ m_2\ u_2)$ be two triangular fuzzy numbers and $(l_1\ m_1\ n_1\ u_1)$ be a trapezoidal fuzzy number, than the basic fuzzy arithmetic operations required to extend conventional AHP algorithms to FAHP are tabulated in Table 2.4.

4.7 Controversies Associated with FAHP Methods

Various review studies are available in the literature summarizing FAHP algorithms [58, 55]. However, in the presence of various priority derivation techniques, none of the review studies on FAHP provide a performance analysis similar to the one present for classical AHP techniques (which was summarized in Section 3.6). Therefore, through this research we intend to fill this gap in the literature and conduct a detail performance analysis of popular FAHP algorithms in terms of the accuracy of the weights calculated from FNPCM. Results of this study are presented in Chapter 7.

Although FAHP has experienced exponential growth over past many years and has been applied in various applications, it has also received its fair share of criticism [59]. Saaty himself criticized fuzzifying human preferences, where he argues that human preferences are already fuzzy in nature and further fuzzifying them will lead to wrong results [31]. Therefore, in this study, we will also attempt to investigate the value of adding fuzziness in human preferences through both empirical and numerical studies. Results from this study are presented in Chapter 8.

Chapter 5

Consistency and Compatibility

In this chapter, we will discuss various measures introduced in the literature to evaluate the quality of NPCMs. People are not like robots and various elements including lapses of reason or concentration, states of mind, trembling, rounding effects and computational mistakes may lead to inconsistent comparisons. Therefore, consistency check is a critical step in AHP so as to evaluates the quality of the NPCMs, and ensures that the decision maker is consistent while providing pairwise comparisons. Consistency measure originates from the transitivity property of pairwise comparison matrix (PCM) i.e., $a_{ij} \times a_{jk} = a_{ik}$. Basic transitivity rules elaborated in [60] are summarized as follows;

- **Intransitivity:** if $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ki} \geq 1$ $\forall i, j, k$
- **Weak Stochastic Transitivity:** if $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ik} \geq 1$ $\forall i, j, k$
- **Strong Stochastic Transitivity:** if $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ik} \geq max(a_{ij}, a_{jk})$ $\forall i, j, k$
- **Additive Transitivity:** if $a_{ij} \times a_{jk} = a_{ik}$ $\forall i, j, k$

Based on these transitivity rules, four main types of inconsistency in PCMs are defined as contradictions, weak contradictions, the numerical scale finiteness and numerical scale discreteness [36]. For a given matrix matrix $A = (a_{ij})_{3\times 3}$ these four different types of inconsistency are briefly explained below with examples;

Contradictions: If $a_{ij} \geq 1$, $a_{jk} \geq 1$ and $a_{ki} \geq 1$ for $i, j, k = 1, 2, 3$ and $i \neq j \neq k$ then consistency in matrix A is caused by contradictions. Cause of this inconsistency can be attributed to the illogical response provided by the decision maker which causes the resulting PCM to have an intransitivity property.

$$\begin{bmatrix} 1 & 3 & 1/4 \\ 1/3 & 1 & 2 \\ 4 & 1/2 & 1 \end{bmatrix}$$

Weak Contradictions: If $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ik} \geq 1$ and $a_{ik} < max(a_{ij}, a_{jk})$ for $i, j, k = 1, 2, 3$ and $i \neq j \neq k$ then consistency in matrix A is caused by weak contradictions. Cause of this inconsistency can also be attributed to the illogical response provided by the decision maker, however, these illogical responses results in weak stochastic transitivity property in the resulting PCM.

$$\begin{bmatrix} 1 & 4 & 3 \\ 1/4 & 1 & 6 \\ 1/3 & 1/6 & 1 \end{bmatrix}$$

Numerical Scale Finiteness: If $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ik} = 9$ and $a_{ij} \times a_{jk} > 9$ for $i, j, k = 1, 2, 3$ and $i \neq j \neq k$ then consistency in matrix A is caused by numerical scale finiteness and this is associated with limitation of the 1-9 scale. In such PCM there exists strong stochastic transitivity properties.

$$\begin{bmatrix} 1 & 3 & 9 \\ 1/3 & 1 & 6 \\ 1/9 & 1/6 & 1 \end{bmatrix}$$

Numerical Scale Discreteness: If $a_{ij} \geq 1$, $a_{jk} \geq 1$, $a_{ik} \geq max(a_{ij}, a_{jk})$ and $a_{ij} \times a_{jk} \neq a_{ik}$ for $i, j, k = 1, 2, 3$ and $i \neq j \neq k$ then consistency in matrix A is caused by numerical scale discreteness and this is also associated with limitation of the 1-9 scale and such matrices also have strong stochastic transitivity properties.

$$\begin{bmatrix} 1 & 1/5 & 2 \\ 5 & 1 & 7 \\ 1/2 & 1/7 & 1 \end{bmatrix}$$

Above discussion further validates the limitations associated with a fixed 1-9 scale, where finiteness and discreteness of this scale is one of the causes for inconsistencies observed in PCMs. Following, we explain Saaty inconsistency index which is the most widely used measure to calculate consistencies in PCMs.

5.1 Saaty Consistency Measure

A matrix is considered to be fully consistent if and only if transitivity rule holds i.e., $a_{ik} \times a_{kj} = a_{ij}$ for all i, j, k. Note that AHP results are based on subjective comparisons assessed from the experts. Humans are very good at comparing two concepts and providing a preferential ordering. However, they are not that good at associating a score on a particular concept and hence in practice comparison matrices are always inconsistent to some degree. Saaty [40] introduces an approach where the consistency of a matrix can be measured by;

$$C.I. = \frac{\lambda_{max} - n}{n - 1}$$

where λ_{max} is the maximum eigenvalue and n is the size of the matrix. Recall that a totally consistent comparison matrix theoretically has only one eigenvalue which is equals to n. As a result, deviation from this theoretical value is used as an indication of inconsistency whose value is given by consistency ratio (C.R).

$$C.R = \frac{C.I}{R.I}$$

R.I is random index whose values are estimated by randomly generating 500 pairwise comparison matrices of different sizes. If $C.R \leq 0.1$ then the given PCM has a reasonable amount of consistency and is regarded as sufficiently consistent, otherwise if $C.R > 0.1$ then the level of inconsistency is on higher side and PCM should be reformed by consulting the experts again. Over the years this measure of inconsistency has been criticized through counter examples in which some PCMs with contradictory judgments were rated as sufficiently consistent and others reasonable PCMs were rejected [47]. Corresponding examples are provided below;

$$C^1 = \begin{bmatrix} 1 & 6 & 9 \\ 1/6 & 1 & 7 \\ 1/9 & 1/7 & 1 \end{bmatrix}, \quad C^2 = \begin{bmatrix} 1 & 2 & 1/2 & 1 & 4 \\ 1/2 & 1 & 2 & 1/2 & 3 \\ 2 & 1/2 & 1 & 1 & 7 \\ 1 & 2 & 1 & 1 & 7 \\ 1/4 & 1/3 & 1/7 & 1/7 & 1 \end{bmatrix}$$

In matrix C^1, criteria 1 is strongly plus more important than criteria 2 (i.e., $S_{13} \to f(S_{13}) = 6$) and criteria 2 is demonstratively more important than criteria 3 (i.e., $S_{14} \to f(S_{14}) = 7$). It is reasonable for decision maker to say that criteria 1 is extremely more important than criteria 3 (i.e., $S_{16} \to f(S_{16}) = 9$). However, as per Saaty inconsistency index, $C.R$ of matrix C^1 is $0.2323 > 0.1$ and hence considered not sufficiently consistent.

In matrix C^2, criteria 1 is weakly more important than criteria 2 (i.e., $S_9 \to f(S_9) = 2$) and criteria 2 is also weakly more important than criteria 3. However, to say criteria 3 is also weakly more important than criteria 1 is an illogical statement. However, as per Saaty inconsistency index, this matrix has $C.R = 0.0933 < 0.1$ and hence regarded as sufficiently consistent. In what follows, we discuss a novel approach to measure inconsistency which attempts to address issues highlighted in this section.

5.2 Consistency Test based on the Consistency Driven Linguistic Methodology

More recently, a novel two step inconsistency test is proposed based on consistency-driven linguistic methodology [36]. It tests whether given LPCMs are logical or not, and later sets interval numerical scales for LPCMs to provide a measure of inconsistency. This approach is graphically illustrated in Figure 5.1.

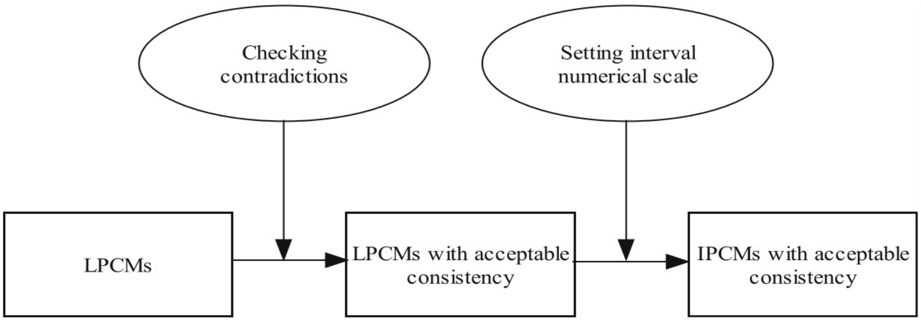


Figure 5.1: Consistency test in consistency driven linguistic methodology

First step is to check whether LPCMs elicited from experts are sufficiently consistent or not. This test is performed by analyzing if given LPCM satisfies strong stochastic transitivity property. An example of sufficiently consistent (Matrix A^1) and inconsistent (Matrix A^2) matrices are as follows;

$$A^1 = \begin{bmatrix} S_8 & S_9 & S_{13} \\ S_7 & S_8 & S_{11} \\ S_3 & S_5 & S_8 \end{bmatrix}, \quad A^2 = \begin{bmatrix} S_8 & S_9 & S_{10} \\ S_7 & S_8 & S_{11} \\ S_6 & S_5 & S_8 \end{bmatrix}$$

Afterwards, a mathematical model is utilized to generate an interval numerical scale which transforms LPCM into Interval Pairwise Comparison Matrices (IPCMs). Due to space limitation, this model is not explained in the proposal report, however, if required its explanation will be provided in subsequent reports. Note that linguistic scale is an ordered set and after the check performed in the first step, LPCM does not contain any contradictions. Corresponding interval scale is also an ordered set and therefore, it does not contain any contradictions as well. Therefore, this measure of inconsistency ensures that no contradiction exists in NPCMs. Furthermore, inconsistencies caused by numerical scale finiteness and discreteness are also addressed accordingly. Such an approach will be useful when testing NPCMs for consistency when personalized scale is used. In this research, we will investigate this novel approach of measuring inconsistency and make relevant proposals.

5.3 Consistency in FAHP

When pairwise comparison matrices are composed of fuzzy numbers, then issue of measuring inconsistency has largely remain sketchy. Although there are numerous consistency measures proposed in the literature to measure consistency from fuzzy pairwise comparison matrices [61] [62] [63] [64] [65] [66], it is observed that most practitioners ignore this critical step of consistency check [67]. Review of the relevant literature shows that there is no universally accepted measure of inconsistency for fuzzy AHP methods. Therefore, we will adopt Saaty measure of inconsistency in our analysis.

5.4 Compatibility Index Value (CIV)

A more robust method to measure inconsistency is through a measure called Compatibility Index Value (CIV) [68] which provides a measure of the deviation between NPCM provided by the decision maker and matrix $W = (w_i/w_j)$ constructed from the derived priority vector from the same NPCM. Let $A = (a_{ij})$ be NPCM provided and $W = (w_i/w_j)$ be perfectly consistent matrix constructed from derived priority vector, then CIV is defined as

$$CIV = n^{-2}.e^T A \circ W^T e \tag{5.1}$$

where n is the size of the matrix and $e^T A \circ W^T e$ is the Hadamard product of matrix A and W^T. Note that if A is perfectly consistent matrix than both matrices A and W will be similar and CIV becomes one, else it will have value greater than one. Therefore, once we have an inconsistent NPCM, and algorithm A derives a priority vector resulting in lower value of CIV compared with a priority vector derived from algorithm B, we can conclude that algorithm A is able to more accurately identify the priorities through information provided by the decision maker through inconsistency pairwise comparisons.

Chapter 6

Research Methodology

Our research methodology will be two fold i.e., numerical and empirical. We will use a data set that was generated numerically as well as gathered through empirical studies to provide answers to the stated research questions. Following we provide a sketch of numerical and empirical study that will be conducted in this research.

6.1 Numerical Study

For the numerical study, we generate a data set of pairwise comparison matrices. Assume that there are n criteria and $w_1, w_2, \cdots, w_n$ are the subjective weights corresponding to these criteria for a decision maker. In practice the decision makers are asked to make pairwise comparisons and it is assumed that they utilize these weights in order to make the comparisons. In reality, inconsistency as well as fuzziness associated with the natural language are incorporated in the process and the resulting matrix would differ from a *theoretically* fully consistent comparison matrix.

In order to conduct the experimental analysis and determine the performance of the nine FAHP algorithms under various conditions, we developed a methodology that mimics the process used by a human decision maker. We generated a random weight vector in each replication and using these subjective weights, we construct a perfectly consistent crisp comparison matrix using Equation 3.3. Afterwards, we added inconsistency into the matrices and fuzzified it through an approach elaborated below.

Procedure to Add Inconsistency: Various levels of inconsistency are added in the matrices through inconsistency parameter $\beta \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. Inconsistency interval $[a\ b]$ is generated for each element of the comparison matrix such that $a = w_i/w_j - \beta \times w_i/w_j$ and $b = w_i/w_j + \beta \times w_i/w_j$. A number x_{ij} is randomly selected from this interval and replaced with the corresponding element of the comparison matrix. Note that while altering matrices in such a manner, the reciprocal nature of the matrices is preserved. Six different values of β are used to generate inconsistency intervals and add desired level of inconsistency in the matrices. However, due to the randomness inherent in the process employed to incorporate inconsistency, it is possible to end up with a comparison matrix which is less or more inconsistent than aimed with the corresponding β parameter. Therefore, once inconsistency is added to the elements of the comparison matrix, the resulting measure of inconsistency is calculated through $C.R = CI/RI$ where $CI = \frac{\lambda_{max}-n}{n-1}$ as suggested by Saaty [40], where λ_{max} is maximum eigenvalue of the comparison matrix and RI is the random index. This C.R measure is employed to classify matrices on different levels of inconsistency i.e., if C.R is between $0 - 0.03$, the corresponding comparison matrix is considered as a low level inconsistent matrix; if C.R is between $0.03-0.06$, it is considered as medium level inconsistent matrix and a comparison matrix with C.R between $0.06 - 0.1$ is regarded as highly inconsistent matrix. Note that any comparison matrix with C.R $\geq$0.1 is considered as a matrix not sufficiently consistent and discarded from the data set as suggested by Saaty [40].

To sum up the above discussion, although parameter β is utilized to add inconsistency in the matrices, C.R measure is used to classify matrices on different inconsistency levels.

Fuzzification of Matrices: The final step of synthetic data generation is fuzzification of matrices which is conducted through parameter $\alpha \in [0, 1]$. This step converts a crisp number into a triangular fuzzy number $[l\ m\ u]$ such that $l = x_{ij} - \alpha$, $m = x_{ij}$ and $u = x_{ij} + \alpha$.

We can generate a similar dataset for the case of original AHP by skipping the fuzzification step of the process. This dataset of crisp NPCMs will be used to evaluate the value of fuzzifying human preferences disccussed in Chapter 7. The overall process of generating dataset is illustrated in Figure 6.1 and the pseudo code in provided in Algorithm 1. This process generates random matrices for $2 < n \leq 15$, representing comparison matrices assessed from decision makers with inconsistency (β) and fuzziness (α).

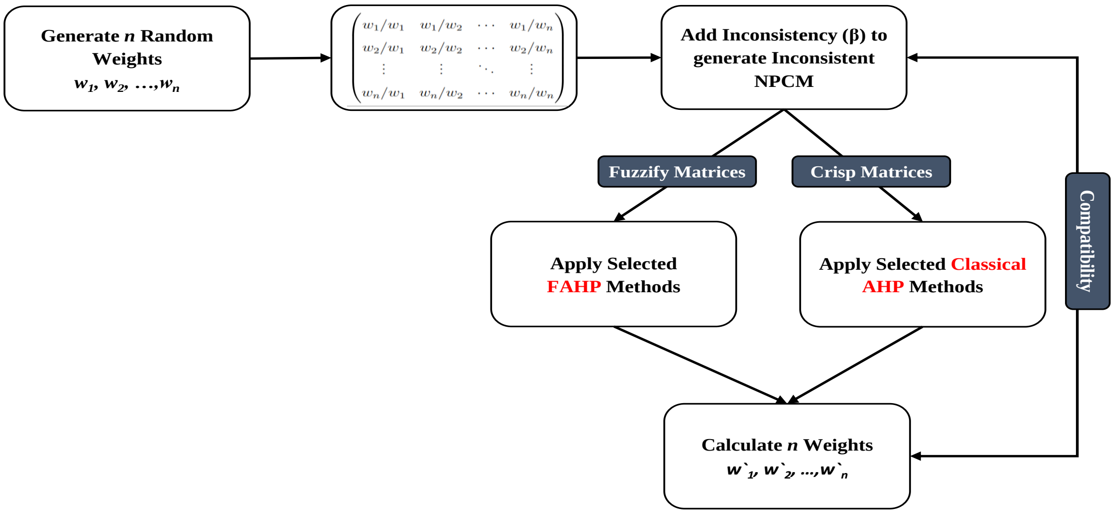


Figure 6.1: Process Diagram

Algorithm 1 Generate Dataset of Pairwise Comparison Matrices

n=*3*
while $n \leq 15$ **do**
    Initial Weights $\leftarrow$ *Generate n Random Weight which sums upto 1*
    Perfectly Consistent Matrix $\leftarrow w_i/w_j$
    $\beta = 0$
    while $\beta \leq 0.4$ **do**
        Add Inconsistency
        $alpha = 0.10$
        while $\alpha \leq 1$ **do**
            if $\alpha = 0$ **then**
                Priority Vector $\leftarrow$ Classical AHP
            else
                $[l\ m\ u][\leftarrow [x_{ij} - \alpha,\ x_{ij},\ x_{ij} + \alpha]$
            Priority Vector $\leftarrow$ Fuzzy AHP
            $\alpha = \alpha + 0.10$
        $\beta = \beta + 0.2$
    $n = n + 4$
Compatibility Index Value (CIV)

6.2 Empirical Study

In addition to numerical study, we also conducted two empirical studies in which pairwise comparisons are elicited from the participants in the form of linguistic labels. Theoretical foundations on the validity of seeking pairwise comparisons is found in the famous theory of scales of measurements proposed by S.S. Stevens [69]. When experts are assessing ratio judgments, in essential they compute ratios in their mind and provide pairwise comparisons accordingly [70]. Over the past decade, mathematical psychologist have attempted to provide further mathematical foundations to understand structural assumptions inherent to the ratio scaling method [71] [72] [73].

One of the ways to validate scientific theory is to show that results from an empirical study match those predicted by the theory itself. Various AHP validation examples are provided in [74] in which through pairwise comparisons, participants were able to accurately predict relative sizes of geometric shapes, relative weights of objects, relative electric consumption of house hold appliances and even relevant wealth of seven nations. Another study conducted similar experiments in which participants provided ratios of distance of pairs of Italian cities from a reference city, ratio of probabilities resulting from games of chance and ratio of rainfall in pairs of European cities, and data acquired through this experiment was used to proposed a novel approach to estimate priority vector through polynomial approximation method [75].

Similar to these studies, we will conduct experiments and seek such pairwise comparisons for which true weights can be measured and already known. An example of such an experiment is illustrated in Figure 6.2 where participants will be shown two images at a time and will be asked to provide pairwise comparisons in the form of linguistic labels such as *"Extremely Dense"*, *"Moderately More Dense"* etc.,.

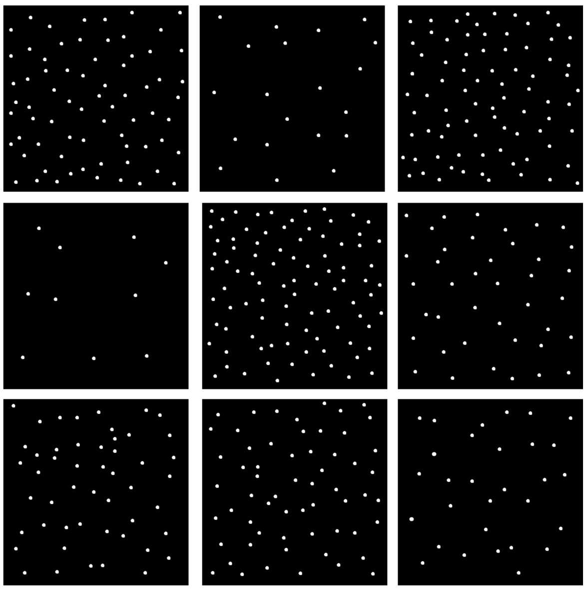


Figure 6.2: Visual Experiment to seek pairwise comparisons of different densities

The participants in both studies were undergraduate students enrolled in Sabanci University, Istanbul, Turkey. In the visual experiment, there were 164 participants; in the mass experiment, 154 participants. For their participation in the experiments, students received 2 bonus points to their course grade. To ensure their focus during the experiments, an additional 1 bonus point was given to the top performing participant in terms of their consistency index value.

Both experiments were approved by the university ethics committee and written consent of the students was obtained prior to participation in the study. Average time required for the visual experiment and mass experiment were approximately 15 and 20 minutes per participant respectively. Since both experiments had an underlying natural scale (i.e., for the visual experiment, the number of dots; for mass experiment, grams), it was possible to assess the *true weights* and corresponding weight vector for both experiments as tabulated in Table 6.1.

Table 6.1: Normalized true weight vector for visual and mass experiment

Number of Dots	Mass of Bottles (Grams)	Weight Vector
10	50	0.0222
20	100	0.0444
30	150	0.0667
40	200	0.0889
50	250	0.1111
60	300	0.1333
70	350	0.1556
80	400	0.1778
90	450	0.2000

6.3 Analysis Methodology for Numerical and Empirical Study

In order to determine the effect of the various experimental parameters, we employed one-way ANOVA (also referred to as a between-subjects ANOVA or one-factor ANOVA) which will help in determining any statistical differences between the mean differences of CIV while we change the experimental parameters. Note that one-way ANOVA is an omnibus test statistic and therefore cannot tell which specific groups of data are significantly different from each other; rather it just provide information that at least two of the groups are significantly different from each other. Therefore, for detailed analysis, we also conducted a post hoc test in order to analyze the results in more detail.

However, in order to conduct one-way ANOVA the following six assumptions must be satisfied:

1. Dependent variables are continuous.

2. Independent variable consists of two or more categorical, independent groups.

3. There is no relationship between the observations in each group or between the groups themselves i.e., independence of observations must hold.

4. There should be no significant outliers, which might have a negative effect on the one-way ANOVA, thus reducing the validity of the results.

5. Dependent variable should be approximately normally distributed for each category of the independent variable. However, one-way ANOVA only requires approximately normal data because it is quite "robust" to violations of normality, meaning that assumption can be a little violated and still provide valid results.

6. Homogeneity of variances must hold.

As explained in the previous section, the structure of the data obtained from the experimental analysis ensures that the first three assumptions are satisfied. We employed the methodology provided in [76] in order to check the latter three assumptions. In order to check the validity of the fourth assumption box-plots are utilized and some outliers are determined. These outliers were neither result of data entry errors nor due to measurement errors but determined to be genuinely unusual values. There are various ways through which these outliers can be treated [77]. We resolve this problem by conducting the analysis with and without these outliers and no significant difference in the results were observed. So we decided to keep these values in our analysis. In order to check the validity of the fifth assumption we conducted both Kolmogorov-Smirnov test as well as Shapiro-Wilk test for each subgroup and determined that our data is normally distributed for each sub group. To test the final assumption, we conducted Levene's test for equality of variances. The results suggested that the assumption of homogeneity of variances was violated. Hence, as suggested in [78] we decided to conduct a modified version of one-way ANOVA which is Welch one-way ANOVA in the analysis.

Chapter 7

Performance Analysis of FAHP Methods

7.1 Results

In this section, the results of the performance analysis conducted on the selected nine FAHP methods will be discussed. Five of these methods are the most popular FAHP methods in the literature while remaining four are fuzzification of original AHP heuristics which are previously discussed in Chapter 4. FAHP methods included in our performance analysis are listed in Table 7.1.

Table 7.1: Selected FAHP methods

Abbreviation	Method
Chang	Original Fuzzy Extent Analysis (FEA) proposed by Chang (1996)
Wang	Fuzzy Extent Analysis (FEA) with modified normalization proposed by Wang (2008)
Laarhoven	Logarithmic Least Squares Method (LLSM) proposed by Laarhoven (1983)
Boender	Logarithmic Least Squares Method (LLSM) with modified normalization proposed by Boender (1989)
Buckley	Geometric Mean Method propose by Buckley (1985)
FAM	Fuzzy Arithmetic Mean similar to arithmetic mean method in original AHP
FGM	Fuzzy Geometric Mean similar to geometric mean method in original AHP
FRSM	Fuzzy Row Sum Method which is Fuzzy Extent Analysis (FEA) with centroid defuzzification
FICSM	Fuzzy Inverse of Column Sum Method

7.1.1 Comparison of Selected nine FAHP methods

We first tabulate descriptive statistics in terms of Mean CIV for selected nine FAHP methods. which shows that mean CIV for FICSM $(1.05968, 0.04665)$ is lowest and for Wang $(1.30748, 0.41495)$ is highest. Same is illustrated in Table 7.2.

Table 7.2: Mean CIV for selected nine FAHP methods

Model	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Boender	624	1.06203	0.04618	0.00185	1.05840	1.06566	1.00003	1.37271
Buckley	624	1.06667	0.04772	0.00191	1.06291	1.07042	1.00003	1.20900
Chang	624	1.23835	0.38388	0.01537	1.20817	1.26853	1.00146	6.48521
FAM	624	1.14855	0.14130	0.00566	1.13744	1.15966	1.00003	2.10831
FGM	624	1.06707	0.04764	0.00191	1.06333	1.07082	1.00003	1.21053
FICSM	624	1.05968	0.04665	0.00187	1.05601	1.06335	1.00003	1.17599
FRSM	624	1.10803	0.09460	0.00379	1.10060	1.11547	1.00006	1.67369
Laarhoven	624	1.06675	0.04763	0.00191	1.06300	1.07049	1.00005	1.32394
Wang	624	1.30748	0.41495	0.01661	1.27486	1.34010	1.00111	6.03032
Total	5616	1.12496	0.21717	0.00290	1.11927	1.13064	1.00003	6.48521

In order to conclude that these differences are significant, one-way Welch ANOVA test is conducted for which results are tabulated in Table 7.3.

Table 7.3: Welch ANOVA analysis between nine different methods

Test	Statistic[a]	df1	df2	Sig.
Welch	84.146	8	2315.467	.000

a. Asymptotically F distributed

Note that Welch ANOVA results only imply that group means differ. It does not show in which particular way these group means differ among various subgroups. Therefore, post hoc test is conducted in order to investigate the significant differences more effectively.

Since the homogeneity of variance assumption is violated, Games-Howell post hoc test is conducted instead of commonly used LSD or Tuckey test. Results of Games-Howell post hoc test corresponding to FICSM vs other FAHP methods are tabulated in Table 7.4. Similar analysis is conducted for each one of the algorithms and results of these analyses are graphically presented in Figure 7.1 as heat map.

Figure 7.1 shows that the mean CIV for FICSM algorithm is lower when compared with other algorithms and these differences are significant for FAM, FRSM, Chang and Wang.

We employed a similar methodology to determine if the above conclusions are also valid under

Table 7.4: Games Howell post hoc test - Comparison of FAHP methods

Method (I)	Method (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
FICSM	Boender	-0.00235	0.00263	0.99327	-0.01052	0.00581
FICSM	Buckley	-0.00699	0.00267	0.18104	-0.01529	0.00131
FICSM	Chang	**-0.17867**	0.01548	0.00000	-0.22685	-0.13049
FICSM	FAM	**-0.08887**	0.00596	0.00000	-0.10740	-0.07034
FICSM	FGM	-0.00740	0.00267	0.12509	-0.01569	0.00090
FICSM	FRSM	**-0.04835**	0.00422	0.00000	-0.06148	-0.03523
FICSM	Laarhoven	-0.00707	0.00267	0.16753	-0.01536	0.00122
FICSM	Wang	**-0.24780**	0.01672	0.00000	-0.29983	-0.19577

**. The mean difference is significant at the 0.05 level.*

Significantly Better	Better	Inferior	Significantly Inferior

	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00464	**-0.17632**	**-0.08652**	-0.00505	0.00235	**-0.04600**	-0.00472	**-0.24545**
Buckley		**-0.17168**	**-0.08188**	-0.00041	0.00699	**-0.04137**	-0.00008	**-0.24081**
Chang			**0.08980**	**0.17127**	**0.17867**	**0.13032**	**0.17160**	-0.06913
FAM				**0.08147**	**0.08887**	**0.04052**	**0.08180**	**-0.15893**
FGM					0.00740	**-0.04096**	0.00033	**-0.24040**
FICSM						**-0.04835**	-0.00707	**-0.24780**
FRSM							**0.04128**	**-0.19945**
Laarhoven								**-0.24073**

Figure 7.1: Heat map - mean CIV differences between nine FAHP methods
**Sample read from the heat map:*
mean CIV of Boender is lower by 0.00464 *as compared to Buckley and this difference is not significant*
mean CIV of Boender is lower by 0.17632 *as compared to Chang and this difference is significant*

various experimental conditions i.e., size of the matrix (n), fuzzification parameter (α) and inconsistency levels ($C.R$) which are discussed below;

7.1.2 Matrix Size

Welch one way ANOVA test (Table 7.5) shows that mean CIV is statistically different for selected four different matrix sizes, $F(3, 3053.866) = 27.866, p < .005$.

Table 7.5: Welch ANOVA analysis between different matrix sizes

Test	Statistic^a	df1	df2	Sig.
Welch	27.866	3	3053.866	.000

a. Asymptotically F distributed.

Games Howell post hoc test is conducted for each FAHP algorithm separately in order to investigate these differences more effectively. Mean CIV differences are illustrated as heat map in Figure 7.2.

Where as I(n) and J(n) refer to mean CIV with different matrix sizes. Result shows that in general, mean CIV is lowest for lower matrix sizes and it increases as the matrix size is increased with the only exception for FEA (Chang and Wang). Otherwise, results are consistent for all other FAHP algorithms as illustrated in Figure 7.3. That is to say, for all FAHP algorithms (except Chang and Wang), mean CIV increases as the matrix size is increased.

Next, we compared performance of selected nine FAHP methods under different matrix sizes. Results are graphically illustrated as heat map in Figure 7.4, which shows that at $n = 3$ the best performing algorithm is Buckley, while at $n = 7$, FICSM is the best performing algorithm. Boender is the best performing algorithm at higher matrix sizes i.e., $n = 11$, $n = 15$. This result will serve as a useful guideline for practitioners in their decision making problems.

		Significantly Better		Better		Inferior		Significantly Inferior		
(I) n	(J) n	Boender	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
3	7	-0.03499	-0.04622	0.03251	-0.09474	-0.04631	-0.03335	-0.06340	-0.03869	0.20329
3	11	-0.04375	-0.05755	-0.10877	-0.16414	-0.05571	-0.04857	-0.10690	-0.04420	0.08642
3	15	-0.04733	-0.06128	-0.05394	-0.21140	-0.05946	-0.05199	-0.14427	-0.04727	0.17501
7	11	-0.00876	-0.01133	-0.14128	-0.06940	-0.00939	-0.01522	-0.04351	-0.00551	-0.11687
7	15	-0.01234	-0.01506	-0.08646	-0.11666	-0.01314	-0.01864	-0.08087	-0.00859	-0.02829
11	15	-0.00358	-0.00373	0.05482	-0.04727	-0.00375	-0.00342	-0.03736	-0.00307	0.08858

Figure 7.2: Post hoc test - Mean CIV differences (I - J) at different matrix sizes for nine FAHP methods

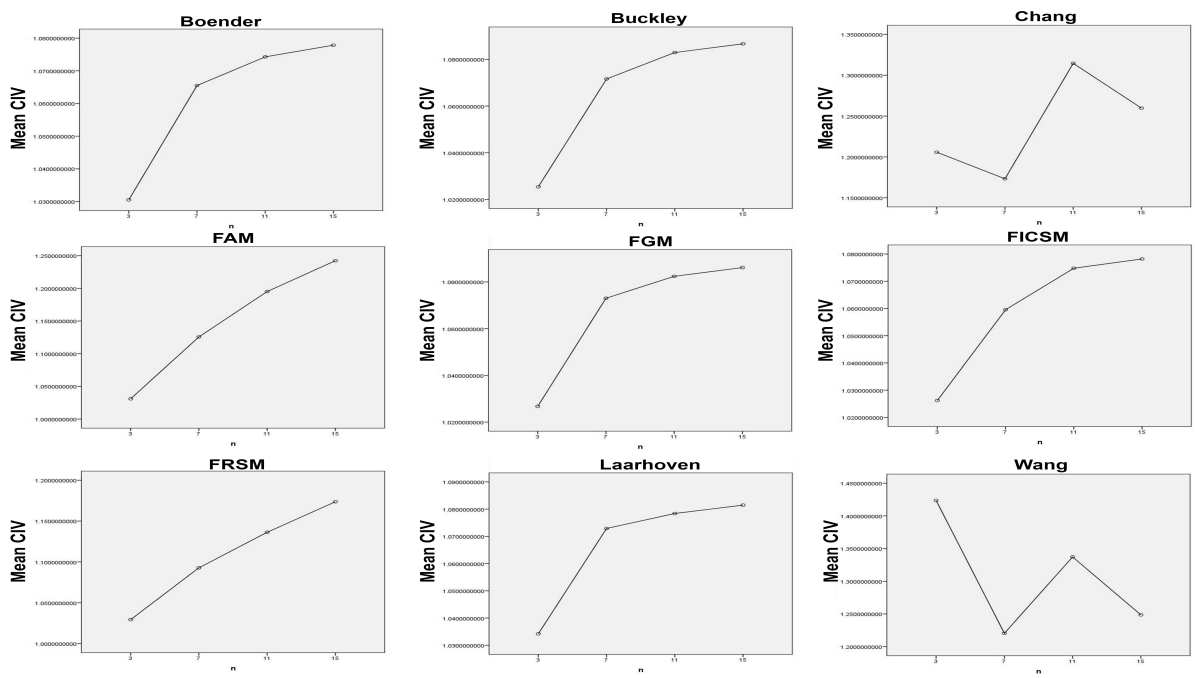


Figure 7.3: Estimated marginal means of CIV

Significantly Better	Better	Inferior	Significantly Inferior

(a)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	0.00511	**-0.17529**	-0.00047	0.00381	0.00431	0.00112	-0.00370	**-0.39315**
Buckley		**-0.18040**	-0.00558	-0.00130	-0.00080	-0.00399	-0.00881	**-0.39826**
Chang			**0.17482**	**0.17909**	**0.17960**	**0.17641**	**0.17159**	**-0.21786**
FAM				0.00428	0.00478	0.00159	-0.00323	**-0.39268**
FGM					0.00050	-0.00269	-0.00750	**-0.39695**
FICSM						-0.00319	-0.00800	**-0.39746**
FRSM							-0.00482	**-0.39427**
Laarhoven								**-0.38945**

(b)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00612	**-0.10778**	**-0.06022**	-0.00752	0.00595	**-0.02729**	-0.00739	**-0.15486**
Buckley		**-0.10166**	**-0.05409**	-0.00140	0.01207	**-0.02116**	-0.00127	**-0.14874**
Chang			0.04757	**0.10027**	**0.11374**	**0.08050**	**0.10039**	-0.04708
FAM				**0.05270**	**0.06617**	**0.03293**	**0.05282**	**-0.09465**
FGM					0.01347	**-0.01977**	0.00012	**-0.14735**
FICSM						**-0.03324**	-0.01335	**-0.16081**
FRSM							**0.01989**	**-0.12758**
Laarhoven								**-0.14747**

(c)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00869	**-0.24030**	**-0.12085**	-0.00815	-0.00051	**-0.06203**	-0.00415	**-0.26297**
Buckley		**-0.23161**	**-0.11216**	0.00054	0.00818	**-0.05334**	0.00455	**-0.25428**
Chang			0.11945	**0.23215**	**0.23979**	**0.17827**	**0.23616**	-0.02267
FAM				**0.11270**	**0.12035**	**0.05882**	**0.11671**	**-0.14212**
FGM					0.00764	-0.05388	0.00400	**-0.25482**
FICSM						-0.06152	-0.00364	**-0.26246**
FRSM							**0.05788**	**-0.20094**
Laarhoven								**-0.25883**

(d)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00884	**-0.18190**	**-0.16454**	-0.00832	-0.00035	**-0.09582**	-0.00364	**-0.17081**
Buckley		**-0.17306**	**-0.15570**	0.00052	0.00850	**-0.08697**	0.00520	**-0.16197**
Chang			0.01736	**0.17358**	**0.18155**	0.08609	**0.17826**	0.01109
FAM				**0.15622**	**0.16419**	**0.06872**	**0.16090**	-0.00627
FGM					0.00797	**-0.08750**	0.00468	**-0.16249**
FICSM						**-0.09547**	-0.00329	**-0.17046**
FRSM							**0.09217**	**-0.07499**
Laarhoven								**-0.16717**

Figure 7.4: Heat map - Mean CIV differences of nine FAHP methods at different matrix sizes (a) n = 3, (b) n = 7, (c) n = 11, (d) n = 15

7.1.3 Fuzzification Parameter

One-way Welch ANOVA test for different levels of fuzzification is tabulated in Table 7.6 which shows that mean CIV is statistically different for four different levels of fuzzification, Welch $F(3, 2936.801) = 8.689, p < .005$.

Table 7.6: Welch ANOVA analysis between different levels of fuzzification

Test	Statistic^a	df1	df2	Sig.
Welch	8.689	3	2936.801	0.000

a. Asymptotically F distributed.

Games Howell post hoc test at different fuzzification levels for nine FAHP models is presented in the form of heat map which is illustrated in Figure 7.5.

Significantly Better	Better	Inferior	Significantly Inferior

(I) alpha	(J) alpha	Boender	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
0.25	0.50	-0.00756	-0.01097	0.12466	**-0.05050**	-0.01115	-0.00372	-0.01876	-0.01041	**0.16503**
0.25	0.75	-0.00778	-0.01008	**0.20704**	**-0.06202**	-0.01042	-0.00405	-0.02443	-0.01073	**0.27610**
0.25	1.00	**-0.02024**	**-0.02841**	**0.23764**	**-0.12255**	**-0.02894**	-0.00883	**-0.06140**	**-0.02914**	**0.35003**
0.50	0.75	-0.00022	0.00089	0.08238	-0.01152	0.00073	-0.00033	-0.00568	-0.00032	0.11107
0.50	1.00	-0.01267	**-0.01744**	0.11299	**-0.07205**	**-0.01779**	-0.00511	**-0.04265**	**-0.01873**	**0.18500**
0.75	1.00	-0.01245	**-0.01833**	0.03061	**-0.06054**	**-0.01852**	-0.00478	**-0.03697**	**-0.01841**	**0.07393**

Figure 7.5: Post hoc test - Mean CIV differences at different levels of α for nine FAHP methods

Figure 7.5 shows that performance of all FAHP algorithms except Chang and Wang decreases as the fuzzification level are increased. This trend is more clearly illustrated in Figure 7.6 which shows that, the performance of FEA (Chang and Wang) increases at higher levels $(0.75, 1.00)$ of fuzzification, whereas, for all other FAHP methods, performance decreases as fuzzification levels are increased. Review of literature on FAHP shows that FEA (Chang and Wang) has been extensively applied in various decision making environments. Therefore, if FEA is the preferred choice of practitioners, then this study recommends to use higher levels of fuzzification in triangular fuzzy numbers.

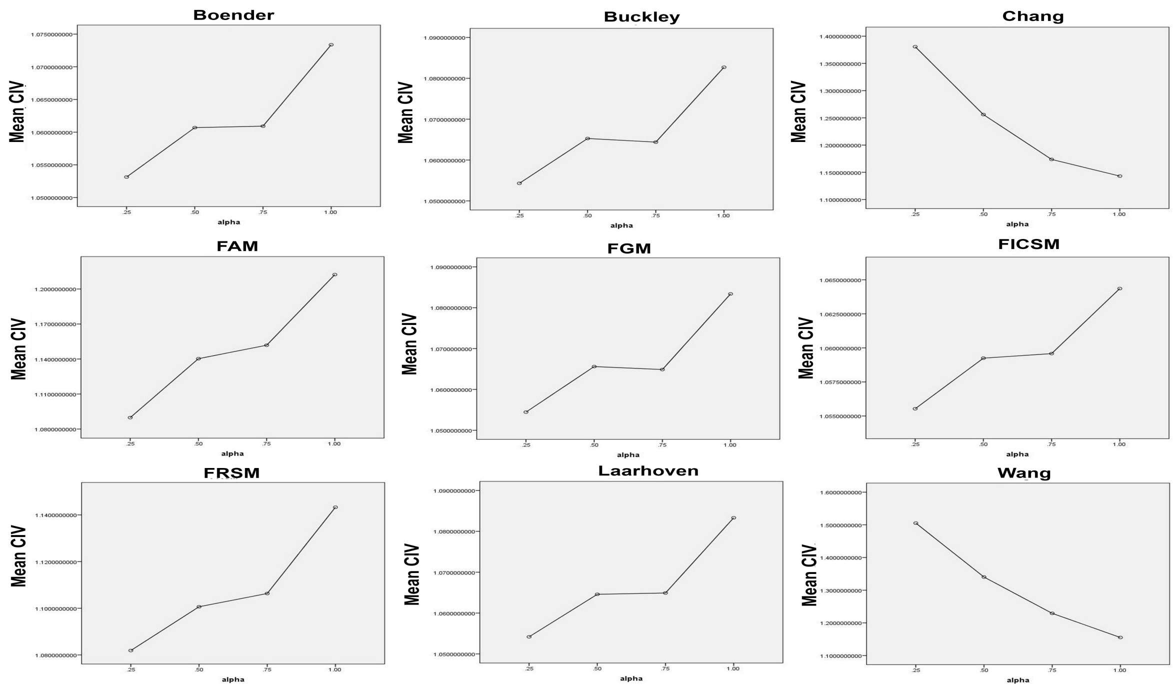


Figure 7.6: Estimated marginal means of CIV

We also conducted comparison of nine FAHP methods under different levels of fuzzification and results are graphically illustrated as a heat map in Figure 7.7 which shows that at lower levels of fuzzification ($\alpha = 0.25$), Boender outperforms all other algorithms. As the fuzzification levels are increased, FICSM is the best performing algorithm. As stated before, performance of FEA (Chang and Wang) seems to improve at higher levels of fuzzification, however, this improved performance is not significant and its performances remains inferior when compared with other

FAHP algorithms.

Significantly Better	Better	Inferior	Significantly Inferior

(a)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00117	**-0.32755**	**-0.03665**	-0.00131	-0.00240	**-0.02875**	-0.00105	**-0.45213**
Buckley		**-0.32638**	**-0.03548**	-0.00014	-0.00123	**-0.02758**	0.00012	**-0.45097**
Chang			**0.29090**	**0.32624**	**0.32515**	**0.29880**	**0.32650**	-0.12459
FAM				**0.03533**	**0.03425**	0.00790	**0.03560**	**-0.41549**
FGM					-0.00108	**-0.02744**	0.00027	**-0.45082**
FICSM						**-0.02635**	0.00135	**-0.44974**
FRSM							**0.02770**	**-0.42338**
Laarhoven								**-0.45109**

(b)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00457	**-0.19533**	**-0.07959**	-0.00490	0.00145	**-0.03994**	-0.00389	**-0.27954**
Buckley		**-0.19076**	**-0.07501**	-0.00033	0.00602	**-0.03537**	0.00068	**-0.27497**
Chang			0.11574	**0.19043**	**0.19678**	**0.15539**	**0.19144**	-0.08421
FAM				**0.07468**	**0.08104**	0.03964	**0.07569**	**-0.19996**
FGM					0.00635	**-0.03504**	0.00101	**-0.27464**
FICSM						**-0.04139**	-0.00534	**-0.28099**
FRSM							**0.03605**	**-0.23960**
Laarhoven								**-0.27565**

(c)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00346	**-0.11273**	**-0.09088**	-0.00395	0.00134	**-0.04540**	-0.00399	**-0.16825**
Buckley		**-0.10926**	**-0.08742**	-0.00049	0.00480	**-0.04194**	-0.00053	**-0.16479**
Chang			0.02185	**0.10878**	**0.11407**	0.06733	**0.10874**	-0.05552
FAM				**0.08693**	**0.09222**	**0.04548**	**0.08689**	-0.07737
FGM					0.00529	**-0.04145**	-0.00004	**-0.16430**
FICSM						**-0.04674**	-0.00533	**-0.16959**
FRSM							**0.04141**	**-0.12285**
Laarhoven								**-0.16426**

(d)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00934	**-0.06967**	**-0.13896**	-0.01002	0.00901	**-0.06992**	-0.00995	**-0.08187**
Buckley		**-0.06033**	**-0.12962**	-0.00068	**0.01835**	**-0.06058**	-0.00061	**-0.07253**
Chang			**-0.06930**	**0.05965**	**0.07868**	-0.00025	**0.05972**	-0.01220
FAM				**0.12895**	**0.14798**	**0.06905**	**0.12902**	0.05710
FGM					**0.01903**	**-0.05990**	0.00007	**-0.07185**
FICSM						**-0.07893**	**-0.01896**	**-0.09088**
FRSM							**0.05997**	-0.01195
Laarhoven								**-0.07192**

Figure 7.7: Heat map - Mean CIV differences of nine FAHP methods at different fuzzification levels
(a) $\alpha = 0.25$, (b) $\alpha = 0.50$, (c) $\alpha = 0.75$, (d) $\alpha = 1.00$

7.1.4 Inconsistency Levels

One-way Welch ANOVA test is conducted for various levels of consistency and results are tabulated in Table 7.7 which shows that mean CIV is statistically different for three levels of inconsistency level, Welch $F(2, 3663.228) = 92.232, p < .005$.

Table 7.7: Welch ANOVA analysis between different levels of inconsistency

Test	Statistic[a]	df1	df2	Sig.
Welch	92.232	2	3663.228	0.000

a. Asymptotically F distributed.

Results from Games Howell post hoc test for different levels of inconsistency for all FAHP methods is graphically illustrated as heat map in Figure 7.8 which shows that performance of all algorithms is higher at lower inconsistency levels as expected. Same is illustrated in Figure 7.9

		Significantly Better		Better		Inferior		Significantly Inferior		
(I) CR	(J) CR	Boender	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Low	Medium	-0.03664	-0.03511	-0.01550	-0.05118	-0.03518	-0.03979	-0.04684	-0.03532	-0.03991
Low	High	-0.08276	-0.08076	-0.14290	-0.11336	-0.08038	-0.09095	-0.10008	-0.08120	-0.12882
Medium	High	-0.04611	-0.04565	-0.12740	-0.06217	-0.04520	-0.05115	-0.05325	-0.04587	-0.08890

Figure 7.8: Post hoc test - Mean CIV differences at different levels of inconsistency (C.R) for nine FAHP methods

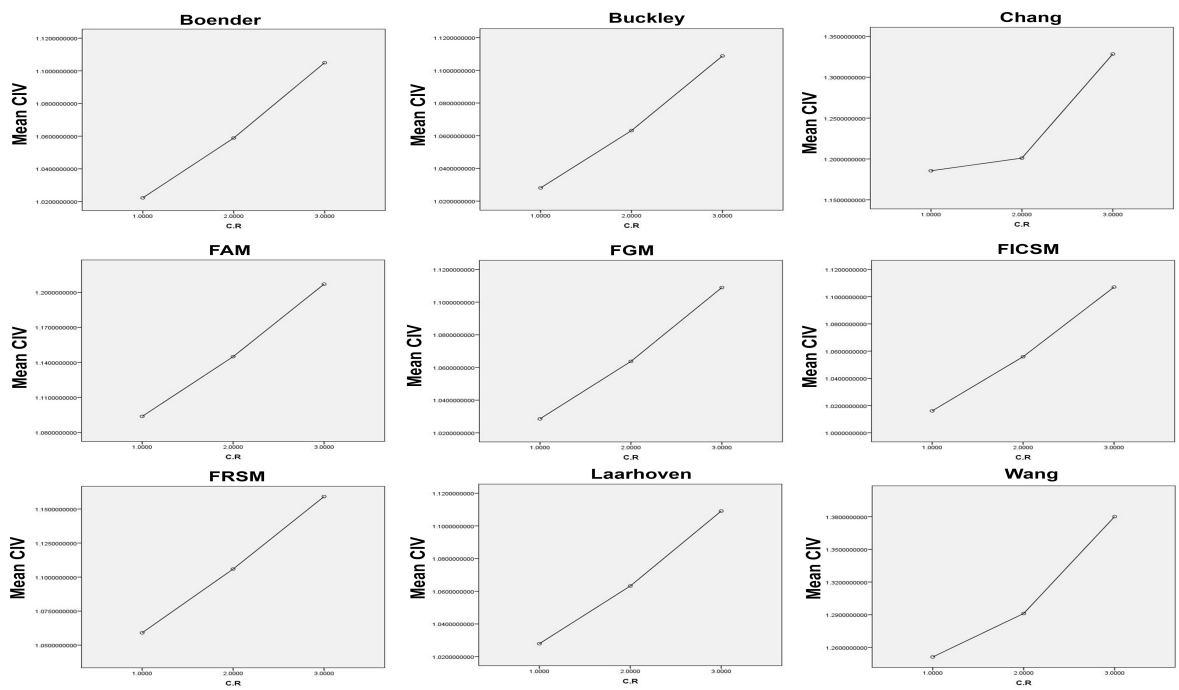


Figure 7.9: Estimated marginal means of CIV

Furthermore, performance of all FAHP methods is analyzed at different levels of inconsistency and results are illustrated as heat map in Figure 7.10

Significantly Better	Better	Inferior	Significantly Inferior

(a)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00581	-0.16332	-0.07147	-0.00632	0.00613	-0.03683	-0.00568	-0.22900
Buckley		-0.15751	-0.06566	-0.00051	0.01194	-0.03102	0.00013	-0.22319
Chang			0.09185	0.15699	0.16945	0.12649	0.15764	-0.06569
FAM				0.06515	0.07760	0.03464	0.06579	-0.15753
FGM					0.01245	-0.03051	0.00064	-0.22268
FICSM						-0.04296	-0.01181	-0.23514
FRSM							0.03115	-0.19218
Laarhoven								-0.22333

(b)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00428	-0.14218	-0.08601	-0.00486	0.00298	-0.04702	-0.00436	-0.23227
Buckley		-0.13790	-0.08173	-0.00058	0.00726	-0.04275	-0.00008	-0.22800
Chang			0.05616	0.13731	0.14516	0.09515	0.13782	-0.09010
FAM				0.08115	0.08899	0.03899	0.08165	-0.14626
FGM					0.00784	-0.04216	0.00050	-0.22741
FICSM						-0.05001	-0.00734	-0.23526
FRSM							0.04266	-0.18525
Laarhoven								-0.22792

(c)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00382	-0.22346	-0.10207	-0.00395	-0.00206	-0.05416	-0.00412	-0.27507
Buckley		-0.21964	-0.09825	-0.00013	0.00176	-0.05034	-0.00030	-0.27125
Chang			0.12139	0.21951	0.22140	0.16931	0.21934	-0.05160
FAM				0.09812	0.10001	0.04792	0.09795	-0.17299
FGM					0.00189	-0.05021	-0.00017	-0.27112
FICSM						-0.05210	-0.00206	-0.27301
FRSM							0.05004	-0.22091
Laarhoven								-0.27095

Figure 7.10: Heat map - Heat map - Mean CIV differences of nine FAHP methods at different inconsistency levels
(a) $C.R$ = Low, (b) $C.R$ = Medium, (c) $C.R$ = High

Figure 7.10 shows that at high inconsistency levels, Boender outperforms all other methods while for other levels FICSM is the best performing algorithm.

7.1.5 Overall Analysis for Boender and FICSM

From the above statistical analysis, it is observed that Boender and FICSM algorithms performed significantly better compared to the other FAHP methods under specific experimental conditions. Here, we present an overall analysis of these two best performing algorithms without considering any experimental condition and using the whole dataset of 624 matrices. The results tabulated in Table 7.8 represent the percentage of matrices, where Boender and FICSM algorithm outperform other FAHP methods in the experiments.

Table 7.8: Overall Analysis

Method	Percentage of matrices in which Boender outperforms the corresponding FAHP algorithm	Percentage of matrices in which FICSM outperforms the corresponding FAHP algorithm
Boender	-------------	48%
Buckley	85%	63%
Chang	87%	91%
FAM	90%	84%
FGM	85%	63%
FICSM	52%	-------------
FRSM	87%	42%
Wang	89%	93%
Laarhoven	93%	62%

7.2 Discussions

In this research we compared performance of nine FAHP methods among which five FAHP methods are the most popular ones in the literature. Compatibility Index Value (CIV) is used as a performance metric to evaluate all nine FAHP methods. Three experimental conditions are considered as part of the analysis, namely, size of the matrix (n), fuzzification level (α) and inconsistency ($C.R$). For the fuzzification parameter four levels are assumed as 0.25, 0.50, 0.75 and 1.00. The fuzzification parameter is not inherent to the problem that the decision maker is facing but more a decision variable as part of the process. That is to say, the decision analysts can set the fuzzification level and conduct FAHP accordingly. On the other hand the inconsistency parameter refers to the inconsistency of the decision maker and is not a decision variable but depends on the fuzzy comparison matrices elicited from the experts. For the analysis three levels are considered for the inconsistency which is low, medium and high based on the consistency ratio (C.R) values as explained before. Finally, four different matrix sizes are considered which are 3, 7, 11 and 15. Note that one can consider 3 as the representative of small sized problems, 7 and 11 are for medium sized problems and 15 for larger cases.

As a result of this set up total of 48 ($= 4 * 3 * 4$) different experimental conditions are constructed. For each condition 13 replications are created randomly. Hence the total dataset is composed of six hundred and twenty four matrices with varying parameters for size of the matrix, fuzzification levels and inconsistency.

Main conclusions of this study are summarized as follows;

1. All three experimental parameters have significant effect on the mean CIV.

- **Size of the Matrix:** As size of the matrix is increased, mean CIV increases and these results are consistent for all selected nine FAHP algorithms. At $n = 3$ the best performing algorithm is Buckley, while at $n = 7$ FICSM method outperforms all other methods. At higher matrix sizes i.e., $n = 11$, $n = 15$, Boender is the best performing algorithm.
- **Fuzzification:** Performance of all FAHP algorithms except Chang and Wang decreases as the fuzzification level increases. Whereas, the performance of FEA methods (Chang and Wang) improves at higher levels of fuzzification, however this improved performance is still inferior when compared with other FAHP algorithms.
- **Inconsistency:** Overall, as the inconsistency levels are increased, mean CIV increases and this increase is consistent over all selected nine FAHP algorithms. At low and medium level of inconsistency FICSM method outperform other methods whereas at higher inconsistency level, Boender is the best performing algorithm.

2. Among the selected nine FAHP algorithms, Boender and FICSM model performs significantly better than other models over various experimental conditions.
3. FEA methods (Chang and Wang) performed inferior compared to other methods, although this is one of the most frequently used technique in the literature.
4. If it is inevitable to use FEA method due to some reason, we propose that one must avoid low levels of fuzzification in triangular fuzzy numbers, as results shows that the performance of this methodology is significantly inferiors at lower levels of α.

Fuzzy numbers are considered as more realistic representations of the imprecise linguistic variables that are used by the decision makers during the preference elicitation stage of the AHP. As a result, abundant of research is being conducted that utilize FAHP and this study will help consolidate this literature. On the other hand, the value of introducing fuzziness to original AHP is yet to be assessed by an extensive experimental analysis and therefore, in the next Chapter we conduct an analysis to determine the value added by fuzzifying human preferences in the context of pairwise comparions and AHP.

7.3 Implications of Results and Proposed Framework for Researchers and Practitioners

This paper provides useful guidance to researchers and practitioners in their selection process of a particular FAHP technique. LLSM method proposed by Boender [27] and FICSM outperformed other selected FAHP methods under various experimental conditions and therefore should be preferred choice of FAHP method in most real life decision making environments.

For example, a company faces a multi-criteria decision problem in which it is in the process of implementing a cloud computing solution and is currently evaluating various cloud computing service providers based on number of selection criteria such as availability, security, storage capacity, acquisition and transaction cost etc. Such a decision problem will have around 7 to 9 selection criteria and thus size of comparison matrices will be 7×7 to 9×9. Figure 7.4 shows that modified LLSM method proposed by Boender [27] performs significantly better under these condition. and thus should be the preferred choice of FAHP method for such an application.

Also Figure 7.7 and 7.6 shows that all FAHP algorithms excluding FEA methods [29] [79] performs significantly better at lower levels of fuzzification. Thus ideal strategy for the implementation of Fuzzy AHP in such a decision problem will be to choose modified LLSM method proposed by Boender [27] as a preferred choice of FAHP algorithm and low level of fuzzifications as a decision variable to achieve best results.

Review of existing literature on FAHP shows that most practitioners utilize FEA methods in their decision problems. This comparative study shows that although FEA is not one of the best performing algorithms, but its performance slightly increases as the level of fuzzification is increased (Figure 7.7 and 7.6). That is to say, if FEA method is the preferred choice of practitioner or researcher in their decision problem then this framework proposes to use higher levels of fuzzification for better results.

Chapter 8

Value of Fuzzifying Human Preferences

In this chapter, the performance of classical AHP and FAHP methods are assessed under various experimental conditions. The Welch Analysis of Variance (ANOVA) tests and Games Howell post hoc test are used as statistical tests with the results summarized as follows.

8.1 Results from Numerical Study

Before focusing on the various experimental conditions, first the performances of all five classical and fuzzy AHP methods are compared in the entire dataset (i.e., 2400 matrices). According to the overall results LLSM (Crisp) has the lowest mean CIV, while Buckley has the worst mean CIV value in our experiments. These results are tabulated in Table 8.1 and graphically illustrated in Figure 8.1.

Table 8.1: Descriptive Statistics - Mean CIV value for five classical AHP and FAHP methods

Method	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LLSM (Crisp)	2400	1.04823	0.04308	0.00088	1.04651	1.04995	1.00000	1.14618
Eigenvector	2400	1.04879	0.04381	0.00089	1.04704	1.05055	1.00000	1.14824
FLLSM (NLP)	2400	1.05879	0.04456	0.00091	1.05701	1.06057	1.00001	1.18302
FLLSM (Boender)	2400	1.06161	0.04608	0.00094	1.05976	1.06345	1.00001	1.39435
Buckley	2400	1.06513	0.04795	0.00098	1.06321	1.06705	1.00001	1.26330

To determine if these mean CIV differences are significant or not ($P \leq 0.05$), Welch ANOVA test is conducted. The results in Table 8.2 indicate that group means significantly differ from each other.

Post hoc test is conducted to investigate the statistical significance of these differences. Note

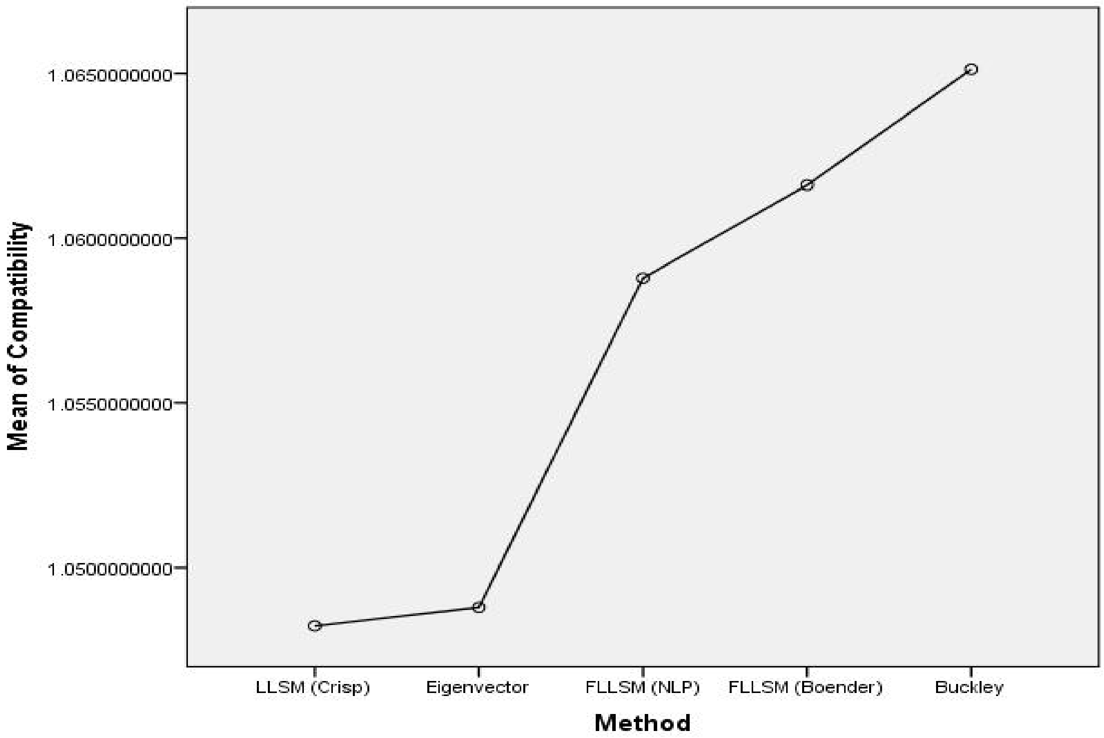


Figure 8.1: Mean CIV for priority vector with PCM

Table 8.2: Welch ANOVA Analysis between nine different methods

Test	Statistic^a	df1	df2	Sig.
Welch	68.705	4	5995.394	.000

a. Asymptotically F distributed

that the homogeneity of variance assumption is violated; therefore, the Games-Howell post hoc test is conducted instead of the commonly used LSD or Tuckey test. Results from Games-Howell post hoc test indicate that all classical AHP methods perform significantly better ($p \leq 0.05$) when compared to FAHP methods in the whole dataset. Figure 8.2 illustrate these differences.

Significantly Better	Better
Inferior	Significantly Inferior

	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.016332	0.003517	0.006337	0.016896
Eigenvector		-0.012815	-0.009996	0.000564
FLLSM (Boender)			0.002819	0.013379
FLLSM (NLP)				0.010560

Figure 8.2: Mean CIV differences for AHP methods

**Sample read from the heat map: the mean CIV of Buckley is higher by* 0.016332 *as compared to Eigenvector and this difference is significant*

As our numerical study consists of various experimental conditions, i.e., size of the matrix (n), fuzzification levels (α) and inconsistency levels (CR), we further investigated the results and evaluated performance of both classical AHP and FAHP methods under these experimental conditions. This discussion is summarized as follows.

Analysis for Different Matrix Sizes: Analyzing results for different matrix sizes demonstrates that as the matrix size increases, performance of all AHP methods (classical and fuzzy) decreases; moreover, this decrease in performance is significant ($p \leq 0.05$). Both classical AHP and FAHP methods perform significantly better at lower matrix sizes, a finding depicted as a heat map in Figure 8.3.

		Significantly Better	Better	Inferior	Significantly Inferior	
(I) n	**(J) n**	**Buckley**	**Eigenvector**	**FLLSM (Boender)**	**FLLSM (NLP)**	**LLSM (Crisp)**
3	7	-0.044419	-0.032822	-0.029575	-0.037058	-0.032101
3	11	-0.054633	-0.044115	-0.036996	-0.046093	-0.043313
3	15	-0.058868	-0.051361	-0.042003	-0.053291	-0.050627
7	11	-0.010215	-0.011293	-0.007421	-0.009035	-0.011212
7	15	-0.014449	-0.018539	-0.012428	-0.016233	-0.018526
11	15	-0.004235	-0.007246	-0.005007	-0.007198	-0.007314

Figure 8.3: Mean CIV differences at different matrix sizes (I)n - (J)n

While comparing different methods at various matrix sizes, the heat map illustrated in Figure 8.4 reveals that for all matrix sizes, classical AHP methods perform significantly better ($p \leq 0.05$) when compared to FAHP methods.

Significantly Better | Better | Inferior | Significantly Inferior

(a)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.008927	-0.008819	0.000967	0.008927
Eigenvector		-0.017746	-0.007960	0.000000
FLLSM (Boender)			0.009786	0.017746
FLLSM (NLP)				0.007960

(b)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.020524	0.006025	0.008328	0.021244
Eigenvector		-0.014499	-0.012196	0.000720
FLLSM (Boender)			0.002304	0.015220
FLLSM (NLP)				0.012916

(c)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.019445	0.008818	0.009508	0.020247
Eigenvector		-0.010627	-0.009938	0.000802
FLLSM (Boender)			0.000690	0.011429
FLLSM (NLP)				0.010739

(d)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.016434	0.008046	0.006544	0.017167
Eigenvector		-0.008388	-0.009890	0.000733
FLLSM (Boender)			-0.001502	0.009121
FLLSM (NLP)				0.010623

Figure 8.4: Games-Howell post hoc test for comparison of AHP methods at different matrix sizes
(a) $n = 3$, (b) $n = 7$, (c) $n = 11$, (d) $n = 15$

Analysis for Different Levels of Fuzzification: Figure 8.5 points out that as the fuzzification levels are increased, performance of all FAHP methods decreases; moreover, this decrease in performance is statistically significant ($p \leq 0.05$). From these results, we can conclude that lower levels of fuzzification should be preferred while applying FAHP methods in a given decision making problem. Such a result entails a considerable doubt on the use of fuzzy scales and/or on the performance of FAHP algorithm that uses these scales

		Significantly Better	Better	Inferior	Significantly Inferior	
(I) α	**(J) α**	**Buckley**	**Eigenvector**	**FLLSM (Boender)**	**FLLSM (NLP)**	**LLSM (Crisp)**
0.10	0.30	**-0.013092**	N/A	**-0.011128**	**-0.008502**	N/A
0.10	0.50	**-0.019442**	N/A	**-0.015694**	**-0.011986**	N/A
0.10	1.00	**-0.032146**	N/A	**-0.024449**	**-0.018771**	N/A
0.30	0.50	-0.006350	N/A	-0.004565	-0.003484	N/A
0.30	1.00	**-0.019054**	N/A	**-0.013321**	**-0.010268**	N/A
0.50	1.00	**-0.012704**	N/A	**-0.008755**	**-0.006785**	N/A

Figure 8.5: Mean CIV differences at different levels of fuzzification (I)α - (J)α

Figure 8.6 reveals that at $\alpha = 0.1$, all classical AHP and FAHP methods perform similarly and there is no statistical significance between the mean CIV differences. That is to say, as the scale used to transform the linguistic labels to numerical values becomes less fuzzy (i.e., more crisp), the algorithms tend to perform similarly. This result verifies the code used during the analysis and also indicates that for the extreme case (no fuzziness in the data), the FAHP methods perform similar to those of classical counterparts. However, as fuzzification levels increase, classical AHP methods start performing better and this improved performance is statistically significant ($p \leq 0.05$). This observation substantiates the doubt regarding the use of fuzzy scales or the FAHP methods as mentioned earlier.

Analysis for Different Levels of Inconsistency: Performance of both classical AHP and FAHP methods decreases as the inconsistency levels are increased (Figure 8.7) and this decrease in performance is statistically significant ($P \leq 0.05$).

The heat map shown in Figure 8.8 reveals that at all levels of inconsistency, classical AHP methods outperform FAHP methods and mean CIV differences are statistically significant ($P \leq 0.05$).

To sum up, the numerical study does not support that fuzzification of the human judgements with fuzzy scales and/or the FAHP methods that are employed to derive the priority vectors from the fuzzy NPCMs add value to the process. On the contrary, for all experimental conditions but one, the FAHP methods are significantly outperformed by their classical AHP

Significantly Better	Better
Inferior	Significantly Inferior

(a)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.000162	0.000165	-0.000018	0.000726
Eigenvector		0.000003	-0.000181	0.000564
FLLSM (Boender)			-0.000183	0.000561
FLLSM (NLP)				0.000745

(b)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.013255	0.002129	0.004571	0.013818
Eigenvector		-0.011126	-0.008683	0.000564
FLLSM (Boender)			0.002443	0.011690
FLLSM (NLP)				0.009247

(c)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.019604	0.003913	0.007437	0.020168
Eigenvector		-0.015691	-0.012167	0.000564
FLLSM (Boender)			0.003524	0.016255
FLLSM (NLP)				0.012731

(d)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.032309	0.007862	0.013357	0.032873
Eigenvector		-0.024446	-0.018952	0.000564
FLLSM (Boender)			0.005495	0.025010
FLLSM (NLP)				0.019515

Figure 8.6: Games-Howell post hoc test for comparison of AHP methods at different fuzzification levels
(a) $\alpha = 0.1$, (b) $\alpha = 0.2$, (c) $\alpha = 0.3$, (d) $\alpha = 0.4$

Significantly Better	Better	Inferior	Significantly Inferior

(I) CR	(J) CR	Buckley	Eigen Vector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Low	Medium	-0.036758	-0.038719	-0.036007	-0.037200	-0.038421
Low	High	-0.081393	-0.085457	-0.080412	-0.081675	-0.084102
Medium	High	-0.044635	-0.046737	-0.044405	-0.044475	-0.045681

Figure 8.7: Mean CIV Differences at different levels of Inconsistency (I)CR - (J)CR

Significantly Better	Better
Inferior	Significantly Inferior

(a)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.018341	0.002940	0.006578	0.018354
Eigenvector		-0.015401	-0.011763	0.000013
FLLSM (Boender)			0.003638	0.015414
FLLSM (NLP)				0.011776

(b)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.016379	0.003691	0.006136	0.016691
Eigenvector		-0.012688	-0.010243	0.000311
FLLSM (Boender)			0.002445	0.013000
FLLSM (NLP)				0.010555

(c)	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.014277	0.003921	0.006296	0.015644
Eigenvector		-0.010356	-0.007981	0.001367
FLLSM (Boender)			0.002375	0.011723
FLLSM (NLP)				0.009348

Figure 8.8: Games-Howell post hoc test for comparison of AHP methods at different inconsistency levels
(a) $CR = Low$, (b) $CR = Medium$, (c) $CR = High$

counterparts. The only exceptional experimental condition is when $\alpha = 0.1$, i.e., the lowest level of fuzzification. One can argue that such low level of fuzzification resembles crisp numbers and rather than using low levels of fuzzification, crisp numbers and correspondingly classical AHP can also be preferred.

8.2 Results from Empirical Study

Apart from the numerical study, we conducted two empirical experiments, i.e., *Visual Experiment* and *Mass Experiment*, in order to evaluate the performance of the classical AHP and FAHP methods. In both experiments, the Linguistic Pairwise Comparison Matrices (LPCMs) were elicited from the participants (164 participants for Visual Experiment and 154 participants for Mass Experiment). Afterwards, these matrices were transformed into numerical Pairwise Comparison Matrices (PCMs) using the 1-9 scale so that priority vectors can be derived using classical AHP methods. For the case of FAHP, LPCMs are transformed into fuzzy PCMs using a scale of 1-9 and two fuzzification parameters were used, namely $\alpha = 0.1$ and $\alpha = 1$. For comparison purposes, note that $\alpha = 0.1$ is selected since it was the level of fuzziness where FAHP methods performed best in the numerical study (Figure 8.5) and $\alpha = 1$ is the popular choice in the literature [55]. Those PCMs having inconsistency ≥ 0.1 were discarded from the data set. Thus, the final empirical dataset for *Visual* and *Mass* experiment consist of 146 and 123 PCMs, respectively.

We first conducted experiments at $\alpha = 0.1$ and results indicate that both classical AHP as well as FAHP methods perform similarly. These findings are consistent with the results of the numerical study presented earlier i.e., when low level of fuzziness is used to transform linguistic labels into fuzzy numbers, FAHP methods do not add any value to the process. For the sake of space, we don‘t present the results for $\alpha = 0.1$ and focus only on the results obtained for the case of $\alpha = 1$.

For each experiment, two different analyses were conducted. In the first analysis, similar to the methodology of a numerical study, we measured the deviation of priority vectors from PCMs provided by the participants (i.e., CIV). On the other hand, for the empirical study, we also have true weights which were tabulated earlier in Table 6.1. Therefore, in the second analysis, we measured the deviation of the derived priority vector from the true weights. For each analysis, descriptive statistics are tabulated in Table 8.3 - 8.6 and corresponding heat maps are illustrated in Figure 8.9 - 8.12. As in the numerical study, classical AHP methods significantly outperform all other FAHP methods for both empirical experiments when $\alpha = 1$.

Table 8.3: Descriptive Statistics - Mean CIV between derived priority vector and PCM (Visual Experiment)

Method	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LLSM (Crisp)	146	1.06175	0.02143	0.00177	1.05824	1.06525	1.02003	1.12318
Eigenvector	146	1.06236	0.02181	0.00180	1.05880	1.06593	1.02011	1.12508
FLLSM (NLP)	146	1.08075	0.02368	0.00196	1.07688	1.08462	1.03358	1.15593
FLLSM (Boender)	146	1.08354	0.02513	0.00208	1.07942	1.08765	1.03454	1.15990
Buckley	146	1.10164	0.03293	0.00273	1.09625	1.10703	1.04055	1.18268

Significantly Better	Better
Inferior	Significantly Inferior

	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.039277	0.018105	0.020890	0.039891
Eigenvector		-0.021172	-0.018387	0.000614
FLLSM (Boender)			0.002785	0.021786
FLLSM (NLP)				0.019001

Figure 8.9: Heat Map - Mean CIV differences between derived priority vector and PCM (Visual Experiment)

Table 8.4: Descriptive Statistics - Mean CIV between derived priority vector and true weights (Visual Experiment)

Method	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LLSM (Crisp)	146	1.13778	0.06567	0.00543	1.12704	1.14852	1.00847	1.29958
Eigenvector	146	1.14246	0.07135	0.00590	1.13079	1.15413	1.00861	1.32607
FLLSM (NLP)	146	1.16123	0.07268	0.00602	1.14934	1.17312	1.02586	1.38829
FLLSM (Boender)	146	1.16476	0.07346	0.00608	1.15275	1.17678	1.02618	1.39644
Buckley	146	1.16400	0.07359	0.00609	1.15196	1.17604	1.02705	1.38542

Significantly Better	Better
Inferior	Significantly Inferior

	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.021542	-0.000763	0.002765	0.026222
Eigenvector		-0.022306	-0.018777	0.004680
FLLSM (Boender)			0.003529	0.026985
FLLSM (NLP)				0.023457

Figure 8.10: Heat Map - Mean CIV differences between derived priority vector and true weights (Visual Experiment)

Table 8.5: Descriptive Statistics - Mean CIV between derived priority vector and PCM (Mass Experiment)

Method	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LLSM (Crisp)	123	1.07560	0.02624	0.00237	1.07091	1.08028	1.02517	1.17033
Eigenvector	123	1.07657	0.02681	0.00242	1.07179	1.08136	1.02530	1.17479
FLLSM (NLP)	123	1.09970	0.03124	0.00282	1.09412	1.10527	1.03968	1.22611
FLLSM (Boender)	123	1.10253	0.03403	0.00307	1.09646	1.10860	1.03968	1.24109
Buckley	123	1.12300	0.05151	0.00464	1.11380	1.13219	1.03980	1.40512

Significantly Better	Better
Inferior	Significantly Inferior

	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.046423	0.020464	0.023299	0.047397
Eigenvector		-0.025959	-0.023124	0.000974
FLLSM (Boender)			0.002835	0.026933
FLLSM (NLP)				0.024098

Figure 8.11: Heat Map - Mean CIV differences between derived priority vector and PCM (Mass Experiment)

Table 8.6: Descriptive Statistics - Mean CIV between derived priority vector and true weights (Mass Experiment)

Method	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LLSM (Crisp)	123	1.14044	0.04851	0.00437	1.13178	1.14910	1.04642	1.29325
Eigenvector	123	1.13966	0.05136	0.00463	1.13049	1.14882	1.04252	1.30339
FLLSM (NLP)	123	1.18210	0.07266	0.00655	1.16913	1.19507	1.05313	1.54426
FLLSM (Boender)	123	1.18640	0.07871	0.00710	1.17235	1.20045	1.05321	1.57801
Buckley	123	1.19879	0.10228	0.00922	1.18053	1.21704	1.05413	1.78107

Significantly Better	Better
Inferior	Significantly Inferior

	Eigenvector	FLLSM (Boender)	FLLSM (NLP)	LLSM (Crisp)
Buckley	0.059129	0.012382	0.016689	0.058346
Eigenvector		-0.046747	-0.042440	-0.000783
FLLSM (Boender)			0.004307	0.045964
FLLSM (NLP)				0.041657

Figure 8.12: Heat Map - Mean CIV differences between derived priority vector and true weights (Mass Experiment)

Even though the above results substantiate the finding that FAHP methods do not add *"any"* value to the process, a more detailed analysis of the results suggests that this observation might be a misleading and hasty conclusion. For example, when we check the individual results for both empirical studies and determine number of times classical AHP method outperforms FAHP methods, we obtain the results tabulated in Table 8.7 and Table 8.8.

Although the classical AHP methods significantly outperform the FAHP methods, Table 8.8 shows that when we compare the deviations of the priority vectors from the true weights for each individual separately, the FAHP (Buckley) method in particular provides more compatible

	FLLSM (Boender)	FLLSM (NLP)	Buckley
LLSM (Crisp)	0	0	0
Eigenvector	5	5	4

	FLLSM (Boender)	FLLSM (NLP)	Buckley
LLSM (Crisp)	1	1	1
Eigenvector	5	5	5

Table 8.7: Number of matrices in which LLSM (Crisp) and Eigenvector method outperforms FAHP method while comparing Mean CIV between **PCM** and **derived priority vector**
(a) Visual Experiment (b) Mass Experiment

	FLLSM (Boender)	FLLSM (NLP)	Buckley
LLSM (Crisp)	44	44	73
Eigenvector	55	55	78

	FLLSM (Boender)	FLLSM (NLP)	Buckley
LLSM (Crisp)	23	25	32
Eigenvector	26	27	31

Table 8.8: Number of matrices in which LLSM (Crisp) and Eigenvector method outperforms FAHP method while comparing Mean CIV between **true weights** and **derived priority vector**
(a) Visual Experiment (b) Mass Experiment

(i.e., closer) priority vectors in almost half of the cases in both of the empirical studies. Note that, the source of inconsistency in pairwise comparison process is not only due to the problems associated with the limitations of the scale used and/or the cognition during the verbal communication phase, but also due to the cognition during the assessment (i.e., valuation) phase. The comparison of the derived priority vectors with the true weights incorporates the cognition as well during the assessment phase of the overall pairwise comparison process. According to the results depicted in Table 8.8, for example in the Visual Experiment, out of the 146 individuals that participated the experiment, Buckley outperforms the LLSM (Crisp) method for 73 individuals and outperforms the Eigenvalue approach for 78 individuals. This observation hints that there might be some benefit of fuzzification of the linguistic preferences even though the existing algorithms are far from exploiting those benefits when the overall analysis is taken into consideration.

Chapter 9

Polarization and Non Polarization Heuristics

Based on the analysis of results presented in the previous chapters, we propose two novel heuristics to map linguistic labels to numbers. In the following sections, first we will provide a theoretical background of AHP and pairwise comparisons and later present details of the proposed heuristics.

9.1 Pairwise Comparisons and AHP

In pairwise comparisons, linguistic qualifiers elicited from experts are transformed into numeric numbers and weight vector is estimated by calculating eigenvector of the Numerical Pairwise Comparison Matrices (NPCMs). This approach assumes that decision maker utilizes a weight vector in his/her mind to provide preferences and process of analyzing pairwise comparison matrices estimate the same weight vector. This process is illustrated in Figure 9.1.

That is to say, the true weight vector and weight vector calculated at the end of process illustrated in Figure 9.1 should be identical, which is not the case in reality due to various reasons explained later in this section.

Note that we can use the special structure of pairwise comparison matrices to further extend the process illustrated in Figure 9.1 and estimate priorities more efficiently. From eigenvector of NPCM, a theoretical NPCM (TNPCM) can be constructed using Equation 9.1. By equating

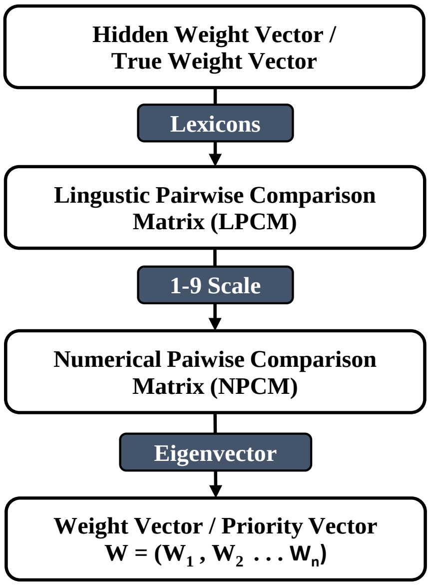


Figure 9.1: Traditional Analytic Hierarchy Process

TNPCM with LPCM provided by the decision maker, quantification of lexicons can be done more efficiently. For clarity this process is illustrated in Figure 9.2a.

$$TNPCM = \begin{pmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{pmatrix} \tag{9.1}$$

While comparing two objects, participants can accurately provide the rank or the order relationship, but the degree of certitude is difficult to capture. Furthermore, in the presence of multiple options with regard to various preference intensities, extreme options are avoided due to extremeness aversion [80] which could lead to inaccurate priority vector. Therefore, using any fixed scale has its inherent shortcomings which may lead to inaccurate weight vector.

In order to address this issue we propose a novel approach which generates a personalized numerical scale during the process and quantify lexicons accordingly. We propose to add polarization phase as shown in Figure 9.2b which convert all preference intensities in to two extreme numbers i.e., 2 and 9 based on a β cut. Details of this process are given in subsection 9.1.1. An argument can be made that decision makers can be given only two extreme options to choose from, however, this could make the decision maker uncomfortable as he/she is forced to choose between two extreme options. Afterwards, personalized numerical scale is generated to construct an individualized pairwise comparison matrix and calculate weight vector accordingly.

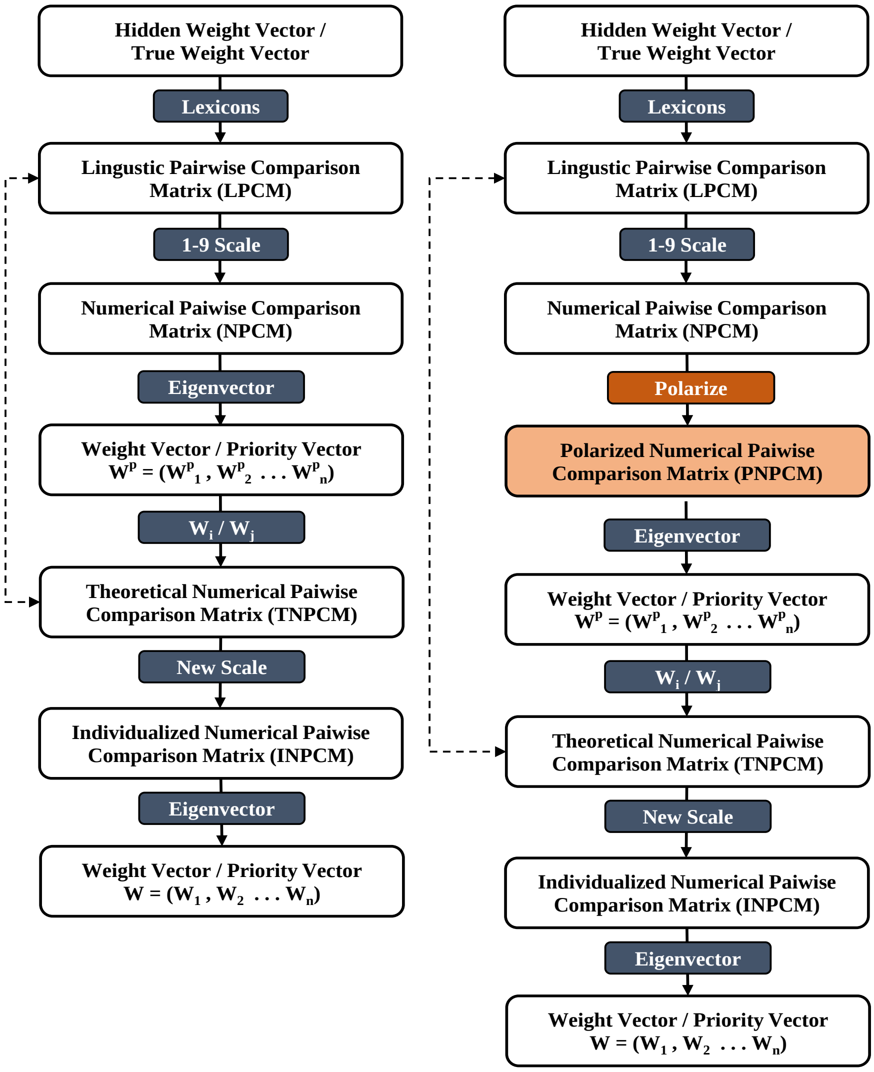


Figure 9.2: (a) Modified Analytic Hierarchy Process (b) Modified Analytic Hierarchy Process with Polarization Heuristics

To construct a heuristic to generate personalized numerical scale, we will use following notations.

Definition 1: Linguistic Pairwise Comparison Matrix (LPCM) Let $S = \{S_i \mid i = 1, 2, ..., 9\}$ represents a linguistic term set such that S_i is defined by a linguistic variable [81]. Let linguistic Pairwise Comparison Matrix (LPCM) represents all pairwise comparisons provided by the decisions maker.

Definition 2: Numerical Pairwise Comparison Matrix (NPCM) Given a numerical scale, function $f()$ transforms linguistic variables into numbers. For example $f^{(Saaty)}$ will transform all linguistic labels into numbers using a scale of 1-9. Furthermore, inverse scale function f^{-1} converts all numeric numbers into corresponding linguistic label.

Definition 3: Consistency of NPCM Let n represent number of criteria, then pairwise comparison matrix is a $n \times n$ real matrix $A = [a_{ij}]$ for all $a_{ij} \geq 0$ and $a_{ij} = 1/a_{ji}$. Let elements of this matrix $[a_{ij}]$ represent preference intensity of i^{th} criteria when compared with j^{th} criteria. Then transitivity relationship, i.e., $a_{ij}.a_{ji} = 1$, provides the measure of inconsistency in the NPCM. Real value of inconsistency is calculated through $CR = CI/RI$ where $CI = \frac{\lambda_{max}-n}{n-1}$ as suggested by Saaty [18], where λ_{max} is a maximum eigenvalue of the comparison matrix and RI is the random index.

Definition 4: Compatibility Index Value Let $A = \{a_{ij}\}$ be NPCM and $W = (w_i/w_j)$ the fully consistent matrix constructed from derived priority vector from A, then CIV is defined as;

$$CIV = n^{-2}.e^T A \circ W^T e \tag{9.2}$$

where n is the size of the matrix and $e^T A \circ W^T e$ is the Hadamard product of matrix A and W^T. Note that if A is a fully consistent matrix then both matrices A and W will be similar and CIV becomes one; otherwise, CIV will have a value greater than one.

9.1.1 Proposed Heuristics

Now we present the heuristic to generate a numerical scale for a particular decision maker.

- **STEP 1:** Let $S = \{S_i \mid i = 1, 2, ..., 9\}$ be a set of available linguistic labels. Construct a linguistic pairwise comparison matrix $LPCM = \{l_{ij}\}$ such that $l_{ij} \in S$.
- **STEP 2:** Let $f^{(Saaty)}$ represent 1-9 numerical scale. Transform matrix L into numerical pairwise comparison matrix $NPCM = \{a_{ij}\}$ and $a_{ij} = f^{(Saaty)}(l_{ij})$
- **STEP 3:** Initiate cut parameter $\beta \in \{3, 4, 5, 6, 7, 8\}$ and iterate the following steps until CIV calculated in Step 10 is minimized.
  - **STEP 4:** Construct a Polarized NPCM matrix $PNPCM = \{p_{ij}\}$ such that *size of* $PNPCM$ = *size of* $NPCM$. If $a_{ij} \leq \beta \rightarrow p_{ij} = 2$ and if $a_{ij} > \beta \rightarrow p_{ij} = 9$. The resulting matrix also holds the reciprocal nature.

- **STEP 5:** Calculate eigenvector of the matrix $PNPCM$ and derive a normalized weight vector w^p.
- **STEP 6:** Construct a perfectly consistent theoretical NPCM ($TNPCM$) using Equation 9.1 from the weight vector w^p calculated in step 5.
- **STEP 7:** Equate matrix $TNPCM$ with linguistic matrix $LPCM$ to estimate numerical values of linguistic labels. In case of multiple values for a specific linguistic label, take average to estimate one single value for a particular linguistic qualifier. The set of numerical values generated in this step is referred to as individualized scale.
- **STEP 8:** Using individualized numerical scale, construct a new individualized NPCM ($INPCM = \{I_{ij}\}$) where $I_{ij} = f^{(Indivualized)}(l_{ij})$.
- **STEP 9:** Calculate eigenvector of matrix $INPCM$ to derive the final prioirty vector.
- **STEP 10:** Calculate CIV between derived priority vector and $INPCM$.

We refer to the heuristic explained above as ***Polarization Heuristics***. In addition to polarization heuristics, we also propose an extension of the original method without polarization as illustrated in Figure 9.2a. We refer to this heuristic as ***Non Polarization Heuristic*** or ***Extended AHP***. We will compare these two methods with the original method proposed by Saaty [18] (Figure 9.1) and approach based on mathematical modeling proposed in [30]. In the following sections, we present results of this comparative study.

9.2 Results and Discussions

In this section we present results for both empirical and numerical studies. As stated before, our main performance metric is Compatibility Index Value (CIV). These results are summarized as follows;

9.2.1 Empirical Results

We first analytically analyze the empirical data and count the number of times a particular linguistic label was given as preference intensity. This data is given in a histogram in Figure 9.3. It shows that fewer number of participants provide strongly worded preferences such as *Extremely More Important*, *Very Very, Strongly More Important*. This points towards the presence of extremeness aversion bias, and supports our claim to polarize the preferences provided by decision makers.

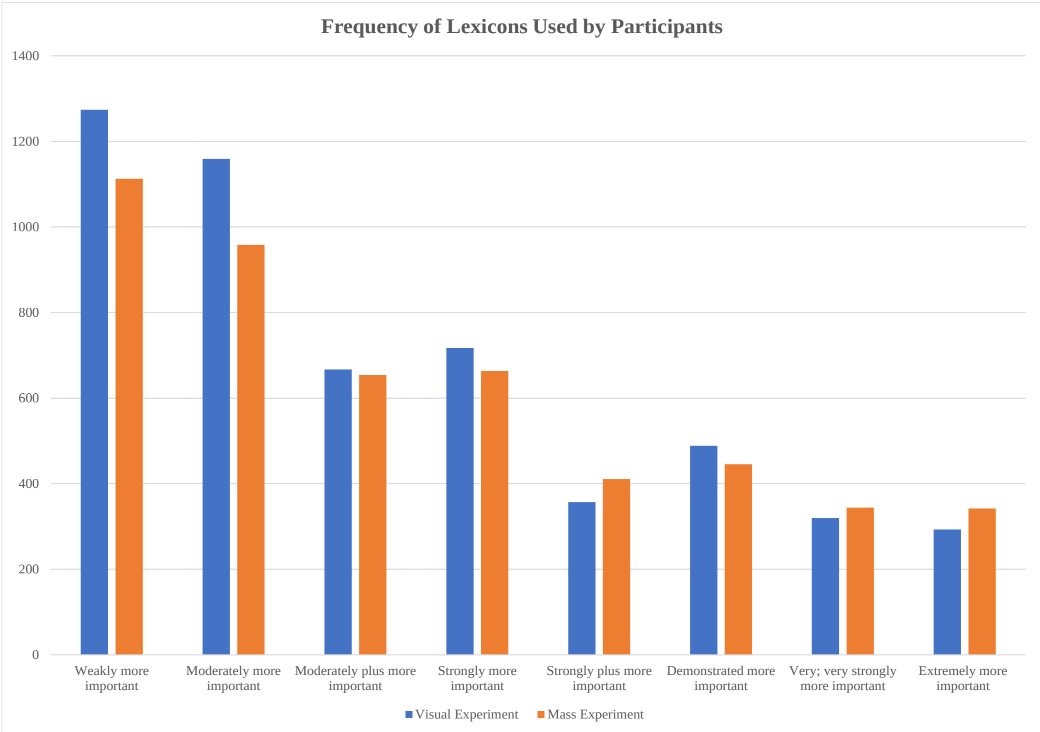


Figure 9.3: Frequency of lexicons used by participants

Next we tabulate mean CIVs and present statistical results. Table 9.1 and Figure 9.4 show that mean CIV between final weight vector and corresponding NPCM for both visual experiment (1.02785) and mass experiment (1.04640) is minimum for LP model. Table 9.2 shows that these differences are statistically significant ($P \leq 0.05$). However, the mean CIV between final weight vector and true weight vector is lowest for polarization heuristics for both visual experiment (1.11001) and mass experiment (1.09178). Table 9.2 shows that these differences are statistically significant. This validates the initial discussion that although LP model due to its structure will yield more consistent matrices, the weight vector calculated from these matrices are found farther from true weight vector.

Table 9.1: Mean CIV for Visual and Mass experiment

Model	**N**	**Mean CIV with NPCM** *(Visual Experiment)*	**Mean CIV with True Weights** *(Visual Experiment)*
1 - 9 Scale	164	1.07727	1.15077
LP Model	164	1.02785	1.12557
Non Polarized	164	1.08163	1.14155
Pol. Heuristics	164	1.05659	1.11001
Model	**N**	**Mean CIV with NPCM** *(Mass Experiment)*	**Mean CIV with True Weights** *(Mass Experiment)*
1 - 9 Scale	154	1.10109	1.14718
LP Model	154	1.04640	1.09973
Non Polarized	154	1.11141	1.11970
Pol. Heuristics	154	1.07520	1.09178

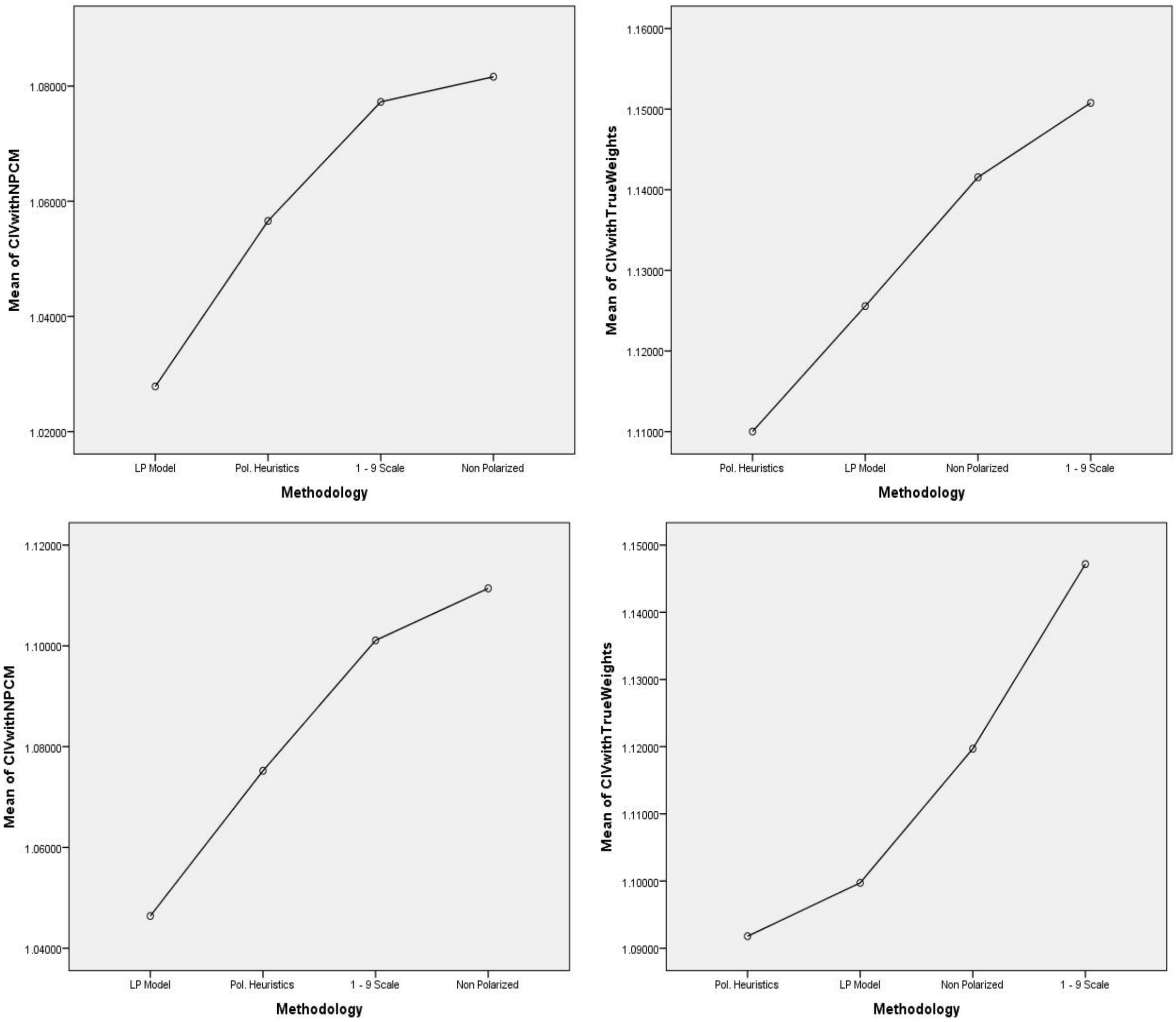


Figure 9.4: Mean CIV Differences

9.2.2 Numerical Results:

Note that while constructing the numerical dataset, we used Saaty scale of 1-9 to convert NPCM into LPCMs. Therefore in our comparison, we exclude this method and compare remaining

Table 9.2: Post Hoc LSD Test for visual and mass experiment

Method	Method	CIV with NPCM *(Visual Experiment)*		CIV with True Weights *(Visual Experiment)*	
(I)	(J)	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.
Pol. Heuristics	LP Model	0.02874	0.000	-0.01556	0.008
Pol. Heuristics	Non Polarized	-0.02504	0.000	-0.03154	0.008
Pol. Heuristics	1 - 9 Scale	-0.02068	0.000	-0.04077	0.008
Non Polarized	LP Model	0.05378	0.000	0.01598	0.008
Non Polarized	1 - 9 Scale	0.00436	0.391	-0.00922	0.008
LP Model	1 - 9 Scale	-0.04942	0.000	-0.02520	0.008
Method	**Method**	**CIV with NPCM *(Mass Experiment)***		**CIV with True Weights *(Mass Experiment)***	
(I)	(J)	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.
Pol. Heuristics	LP Model	0.02880	0.00001	-0.00795	0.144
Pol. Heuristics	Non Polarized	-0.03620	0.00000	-0.02792	0.000
Pol. Heuristics	1 - 9 Scale	-0.02589	0.00005	-0.05540	0.000
Non Polarized	LP Model	0.06501	0.00000	0.01997	0.000
Non Polarized	1 - 9 Scale	0.01032	0.10253	-0.02749	0.000
LP Model	1 - 9 Scale	-0.05469	0.00000	-0.04746	0.000

three methods. Tables 9.3 - 9.5 tabulate mean CIVs for different matrix sizes. Similar to the empirical study, results show that mean CIV between final weight vector and corresponding NPCM is minimum for LP model for all matrix sizes. Results from LSD post hoc test (Table 9.4) show that these differences are statistically significant ($P \leq 0.05$).

NPCMs constructed from LP model are highly consistent and therefore weight vector calculated from these matrices will yield minimum CIV when compared with corresponding NPCM. However, rather than comparing final weight vector with its corresponding NPCM, comparing it with true weight vector will provide more insight as it will provide a measure of deviation of final weight vector from the hidden weight vector which was used by the decision maker during the elicitation stage.

Table 9.3: Mean CIV with NPCM

Model	N	Mean *(n = 3)*	Mean *(n = 7)*	Mean *(n = 11)*	Mean *(n = 15)*
LP Model	225	1.00218	1.00789	1.01048	1.01411
Non Polarized	225	1.01006	1.03900	1.04741	1.05106
Pol. Heuristics	225	1.00794	1.02386	1.02777	1.02893

Table 9.4: Post Hoc LSD Test - CIV with NPCM

Method	Method	n = 3		n = 7		n = 11		n = 15	
(I)	(J)	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.
LP Model	**Pol. Heuristics**	**-0.00576**	0.000	**-0.01597**	0.000	**-0.01729**	0.000	**-0.01482**	0.000
LP Model	**Non Polarized**	**-0.00789**	0.000	**-0.03111**	0.000	**-0.03693**	0.000	**-0.03695**	0.000
Pol. Heuristics	**Non Polarized**	-0.00213	0.053	**-0.01514**	0.000	**-0.01964**	0.000	**-0.02213**	0.000

Table 9.5 shows that mean CIV between final weight vector and true weight vector is minimum

for polarization heuristic at $n = 3$ and results from LSD post hoc test tabulated in Table 9.6 show that these differences are statistically significant ($P \leq 0.05$). For $n = 7, 11, 15$, non polarized heuristic (or extended AHP heuristic) provide the lowest CIV between final weight vector and true weight vector and these differences are statistically significant ($P \leq 0.05$).

Numerical study also shows that LP model yields more consistent matrices, however, in an effort to make NPCMs more consistent, they become farther from the true weight vector, thus adversely affecting the validity of the outcome. Aim of such decision tools should be to estimate the weight vector with least deviation from the hidden weight vector utilized by the decision maker while providing his/her pairwise comparisons.

Table 9.5: Mean CIV with true weights

Model	**N**	**Mean** ***(n = 3)***	**Mean** ***(n = 7)***	**Mean** ***(n = 11)***	**Mean** ***(n = 15)***
LP Model	225	1.07531	1.07959	1.08061	1.08200
Non Polarized	225	1.05703	1.03361	1.03325	1.03562
Pol. Heuristics	225	1.03901	1.05146	1.05526	1.06184

Table 9.6: Post Hoc LSD Test - CIV with true weights

Method	Method	n = 3		n = 7		n = 11		n = 15	
(I)	(J)	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.	Mean Diff. (I-J)	Sig.
LP Model	Pol. Heuristics	**0.03630**	0.000	**0.02813**	0.000	**0.02535**	0.000	**0.02016**	0.000
LP Model	Non Polarized	**0.01828**	0.033	**0.04598**	0.000	**0.04736**	0.000	**0.04638**	0.000
Pol. Heuristics	Non Polarized	**-0.01802**	0.036	**0.01785**	0.000	**0.02201**	0.000	**0.02622**	0.000

Chapter 10

Conclusions and Future Research Areas

10.1 Conclusions

Most decision making environments contain qualitative information in the form of verbal phrases and quantification of such phrases has remained a contentious issue in the literature. Furthermore, as we move towards the fourth industrial revolution and human-machine teaming becomes more frequent, accurate interpretation of such verbal phrases is of critical importance. Analytic Hierarchy Process (AHP) can aid in this process as it elicit human preferences in the form of linguistic labels and utilizes eigenvector of numerical pairwise comparison matrices to estimate weight vector of the available criteria. Even though AHP is among the most commonly used and widely studied operations research techniques, there are still plenty of issues requiring clarification and improvement.

One such issue is the incorporation of fuzzy set theory in the AHP process, where fuzzy numbers are considered by some as more realistic representations of the linguistic comparisons employed by the decision makers during the preference elicitation stage of AHP, while others oppose this idea. Until now, arguments from both sides were merely theoretical and neither numerical nor empirical analysis were conducted to shed light on the issue. To bridge this gap and evaluate the value of FAHP methods, our research compared performances of most popular FAHP methods. Results show that among the selected nine FAHP methods, modified fuzzy logarithmic least squares method proposed by Boender [27] and FICSM proposed in this study performs significantly better than other methods over various experimental conditions.

We also conducted a detailed study to investigate the value of introducing fuzziness to an original AHP. Both numerical and empirical studies suggest that original AHP methods significantly outperformed all the FAHP methods considered in this research. At low levels of fuzzification, FAHP methods yielded similar results with the classical AHP methods; however, low levels of fuzzification can be considered as equivalent to the crisp case and thus the essence of fuzzifying human preferences is lost. Based on the results of the numerical study in particular, one can conclude that FAHP methods presented in the literature fail to assess the hidden priority vectors from the fuzzy pairwise comparison matrices, and thus, do not add any value to the process. On the other hand, as the analysis at the end of Section 5.2 reveals, a hasty conclusion that implies FAHP is totally redundant/irrelevant might also be misleading. Therefore, the results of our research suggest that the existing FAHP methods are far from exploiting any potential benefit of fuzzy representation of the pairwise comparisons in AHP; thus these methods should be avoided until either they are modified or novel algorithms that are validated with this type of studies are developed.

Finally, we utilized the special structure of pairwise comparison matrices and proposed two simple heuristic to estimate linguistic qualifiers more efficiently. Results from empirical and numerical studies show that methods proposed in this research yield priority vector closer to true weights.

10.2 Future Research Areas

In this research, we target lexicons in the domain of MCDM; however, for future research we propose to implement similar framework for lexicons used in other domains. This study can be further extended using artificial intelligence techniques such as neural networks. Furthermore, Ventromedial Prefrontal Cortex (VMF) is an important part of brain which is responsible for human decision making. Using FMRI technology, real time neuron activity in this part of the brain can be measured during the elicitation of pairwise comparisons, which could provide more insight into interpretation of linguistic preferences elicited from the decision makers.

Bibliography

[1] Colin F Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics: Why economics needs brains. *Scandinavian Journal of Economics*, 106(3):555–579, 2004.

[2] Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[3] Graham Loomes. Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, 8(4):301–323, 2005.

[4] Pavlo R Blavatskyy. Stochastic expected utility theory. *Journal of Risk and Uncertainty*, 34(3):259–286, 2007.

[5] Ralph L Keeney and Howard Raiffa. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

[6] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

[7] Graham Loomes and Robert Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824, 1982.

[8] Ben Seymour and Samuel M McClure. Anchors, scales and the relative coding of value in the brain. *Current Opinion in Neurobiology*, 18(2):173–178, 2008.

[9] Robert M Roe, Jermone R Busemeyer, and James T Townsend. Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, 108(2):370, 2001.

[10] Andrew E Clark and Andrew J Oswald. Satisfaction and comparison income. *Journal of Public Economics*, 61(3):359–381, 1996.

[11] Amos Tversky. Intransitivity of preferences. *Psychological Review*, 76(1):31, 1969.

[12] Gerd Gigerenzer. Fast and frugal heuristics: The tools of bounded rationality. *Blackwell Handbook of Judgment and Decision Making*, 62:88, 2004.

[13] Elke U Weber, Eric J Johnson, Kerry F Milch, Hannah Chang, Jeffrey C Brodscholl, and Daniel G Goldstein. Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, 18(6):516–523, 2007.

[14] Valerie F Reyna. A theory of medical decision making and health: fuzzy trace theory. *Medical Decision Making*, 28(6):850–865, 2008.

[15] Thomas S Wallsten, David V Budescu, Rami Zwick, and Steven M Kemp. Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, 31(2):135–138, 1993.

[16] Eelko KRE Huizingh and Hans CJ Vrolijk. A comparison of verbal and numerical judgments in the analytic hierarchy process. *Organizational Behavior and Human Decision Processes*, 70(3):237–247, 1997.

[17] David V Budescu and Thomas S Wallsten. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, 36(3):391–405, 1985.

[18] Thomas L Saaty. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281, 1977.

[19] Jack Davis. Sherman kent and the profession of intelligence analysis. Technical report, CENTRAL INTELLIGENCE AGENCY WASHINGTON DC, 2002.

[20] Eng U Choo, William C Wedley, and Diederik JD Wijnmalen. Mathematical support for the geometric mean when deriving a consistent matrix from a pairwise ratio matrix. *Fundamenta Informaticae*, 144(3-4):263–278, 2016.

[21] Patrick T Harker and Luis G Vargas. The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science*, 33(11):1383–1403, 1987.

[22] FA Lootsma. Conflict resolution via pairwise comparison of concessions. *European Journal of Operational Research*, 40(1):109–116, 1989.

[23] FJ Dodd and HA Donegan. Comparison of prioritization techniques using interhierarchy mappings. *Journal of the Operational Research Society*, 46(4):492–498, 1995.

[24] Ahti A Salo and Raimo P Hämäläinen. On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, 6(6):309–319, 1997.

[25] Alessio Ishizaka, Dieter Balkenborg, and Todd Kaplan. Influence of aggregation and measurement scale on ranking a compromise alternative in ahp. *Journal of the Operational Research Society*, 62(4):700–710, 2011.

[26] PJM Van Laarhoven and Witold Pedrycz. A fuzzy extension of saaty's priority theory. *Fuzzy Sets and Systems*, 11(1):199–227, 1983.

[27] CGE Boender, JG De Graan, and FA Lootsma. Multi-criteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems*, 29(2):133–143, 1989.

[28] James J Buckley. Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*, 17(3):233–247, 1985.

[29] Da-Yong Chang. Applications of the extent analysis method on fuzzy ahp. *European Journal of Operational Research*, 95(3):649–655, 1996.

[30] Yucheng Dong, Wei-Chiang Hong, Yinfeng Xu, and Shui Yu. Numerical scales generated individually for analytic hierarchy process. *European Journal of Operational Research*, 229(3):654–662, 2013.

[31] Thomas L Saaty and Liem T Tran. On the invalidity of fuzzifying numerical judgments in the analytic hierarchy process. *Mathematical and Computer Modelling*, 46(7):962–975, 2007.

[32] Robert M Hamm. Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, 48(2):193–223, 1991.

[33] D Ma and X Zheng. 9/9-9/1 scale method of ahp. In *Proceedings of the 2nd International Symposium on the AHP*, volume 1, pages 197–202, 1991.

[34] Alessio Ishizaka, Dieter Balkenborg, and Todd Kaplan. Influence of aggregation and preference scale on ranking a compromise alternative in ahp. In *Multidisciplinary Workshop on Advances in Preference Handling*, 2006.

[35] Jiří Franek and Aleš Kresta. Judgment scales and consistency measure in ahp. *Procedia Economics and Finance*, 12:164–173, 2014.

[36] Hengjie Zhang, Xin Chen, Yucheng Dong, Weijun Xu, and Shihua Wang. Analyzing saatys consistency test in pairwise comparison method: a perspective based on linguistic and numerical scale. *Soft Computing*, pages 1–11, 2016.

[37] Sheng-Hshiung Tsaur, Te-Yi Chang, and Chang-Hua Yen. The evaluation of airline service quality by fuzzy mcdm. *Tourism Management*, 23(2):107–115, 2002.

[38] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

[39] Gordon Crawford and Cindy Williams. A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4):387–405, 1985.

[40] Thomas L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation (Decision Making Series)*. Mcgraw-Hill (Tx), 1980.

[41] KO Cogger and PL Yu. Eigenweight vectors and least-distance approximation for revealed preference in pairwise weight ratios. *Journal of Optimization Theory and Applications*, 46(4):483–491, 1985.

[42] ATW Chu, RE Kalaba, and K Spingarn. A comparison of two methods for determining the weights of belonging to fuzzy sets. *Journal of Optimization Theory and Applications*, 27(4):531–538, 1979.

[43] N Byson. A goal programming method for generating priorities vectors. *Journal of Operational Research Society, Palgrave Macmillan Ltd., Houndmills, Basingstoke, Hampshire, RG21 6XS, England*, pages 641–648, 1995.

[44] Thomas L Saaty and G Hu. Ranking by eigenvector versus other methods in the analytic hierarchy process. *Applied Mathematics Letters*, 11(4):121–125, 1998.

[45] Thomas L Saaty. Decision-making with the ahp: Why is the principal eigenvector necessary. *European Journal of Operational Research*, 145(1):85–91, 2003.

[46] Eng Ung Choo and William C Wedley. A common framework for deriving preference values from pairwise comparison matrices. *Computers & Operations Research*, 31(6):893–908, 2004.

[47] Carlos A Bana e Costa and Jean-Claude Vansnick. A critical analysis of the eigenvalue method used to derive priorities in ahp. *European Journal of Operational Research*, 187(3):1422–1428, 2008.

[48] B Golany and M Kress. A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices. *European Journal of Operational Research*, 69(2):210–220, 1993.

[49] Alessio Ishizaka and Markus Lusti. How to derive priorities in ahp: a comparative study. *Central European Journal of Operations Research*, 14(4):387–400, 2006.

[50] David V Budescu, Rami Zwick, and Amnon Rapoport. A comparison of the eigenvalue method and the geometric mean procedure for ratio scaling. *Applied Psychological Measurement*, 10(1):69–78, 1986.

[51] Thomas L Saaty and Luis G Vargas. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. *Mathematical Modelling*, 5(5):309–324, 1984.

[52] E Takeda, KO Cogger, and PL Yu. Estimating criterion weights using eigenvectors: A comparative study. *European Journal of Operational Research*, 29(3):360–369, 1987.

[53] Fatemeh Zahedi. A simulation study of estimation methods in the analytic hierarchy process. *Socio-Economic Planning Sciences*, 20(6):347–354, 1986.

[54] Nurul Adzlyana Mohd Saadon, Rosma Mohd Dom, and Daud Mohamad. Comparative analysis of criteria weight determination in ahp models. In *Science and Social Research (CSSR), 2010 International Conference on*, pages 965–969. IEEE, 2010.

[55] Gülçin Büyüközkan, Cengiz Kahraman, and Da Ruan. A fuzzy multi-criteria decision approach for software development strategy selection. *International Journal of General Systems*, 33(2-3):259–280, 2004.

[56] Ying-Ming Wang, Taha Elhag, and Zhongsheng Hua. A modified fuzzy logarithmic least squares method for fuzzy analytic hierarchy process. *Fuzzy Sets and Systems*, 157(23):3055–3071, 2006.

[57] Ying-Ming Wang and Taha M.S. Elhag. On the normalization of interval and fuzzy weights. *Fuzzy Sets and Systems*, 157(18):2456 – 2471, 2006.

[58] Mohammad Ataei, Reza Mikaeil, Seyed Hadi Hoseinie, and Seyed Mehdi Hosseini. Fuzzy analytical hierarchy process approach for ranking the sawability of carbonate rock. *International Journal of Rock Mechanics and Mining Sciences*, 50:83–93, 2012.

[59] Kèyù Zhü. Fuzzy analytic hierarchy process: Fallacy of the popular methods. *European Journal of Operational Research*, 236(1):209–217, 2014.

[60] Enrique Herrera-Viedma, Francisco Herrera, Francisco Chiclana, and María Luque. Some issues on consistency of fuzzy preference relations. *European Journal of Operational Research*, 154(1):98–109, 2004.

[61] Lawrence C Leung and D Cao. On consistency and ranking of alternatives in fuzzy ahp. *European Journal of Operational Research*, 124(1):102–113, 2000.

[62] Ying-Ming Wang, Jian-Bo Yang, and Dong-Ling Xu. Interval weight generation approaches based on consistency test and interval comparison matrices. *Applied Mathematics and Computation*, 167(1):252–273, 2005.

[63] Rafikul Islam, MP Biswal, and SS Alam. Preference programming and inconsistent interval judgments. *European Journal of Operational Research*, 97(1):53–62, 1997.

[64] Jaroslav Ramík and Petr Korviny. Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems*, 161(11):1604–1613, 2010.

[65] Yejun Xu, Xia Liu, and Huimin Wang. The additive consistency measure of fuzzy reciprocal preference relations. *International Journal of Machine Learning and Cybernetics*, pages 1–12, 2017.

[66] Wu-E Yang, Chao-Qun Ma, Zhi-Qiu Han, and Wen-Jun Chen. Checking and adjusting order-consistency of linguistic pairwise comparison matrices for getting transitive preference relations. *OR Spectrum*, 38(3):769–787, 2016.

[67] Bülent Başaran. A critique on the consistency ratios of some selected articles regarding fuzzy ahp and sustainability. *3rd International Symposium on Sustainable Development*, 2012.

[68] TL Saaty. A ratio scale metric and the compatibility of ratio scales: The possibility of arrow's impossibility theorem. *Applied Mathematics Letters*, 7(6):51–57, 1994.

[69] Stanley Smith Stevens et al. On the theory of scales of measurement. *On the Theory of Scales of Measurement*, 1946.

[70] Louis Narens. A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, 40(2):109–129, 1996.

[71] Louis Narens. The irony of measurement by subjective estimations. *Journal of Mathematical Psychology*, 46(6):769–788, 2002.

[72] R Duncan Luce. A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, 109(3):520, 2002.

[73] R Duncan Luce. Symmetric and asymmetric matching of joint presentations. *Psychological Review*, 111(2):446, 2004.

[74] Rozann Whitaker. Validation examples of the analytic hierarchy process and analytic network process. *Mathematical and Computer Modelling*, 46(7-8):840–859, 2007.

[75] Michele Bernasconi, Christine Choirat, and Raffaello Seri. The analytic hierarchy process and the theory of measurement. *Management Science*, 56(4):699–711, 2010.

[76] Glenn Gamst, Lawrence S Meyers, and AJ Guarino. *Analysis of variance designs: A conceptual and computational approach with SPSS and SAS*. Cambridge University Press, 2008.

[77] Jason W Osborne and Amy Overbay. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, 9(6):1–12, 2004.

[78] Lisa M Lix, Joanne C Keselman, and HJ Keselman. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test. *Review of Educational Research*, 66(4):579–619, 1996.

[79] Ying-Ming Wang, Ying Luo, and Zhongsheng Hua. On the extent analysis method for fuzzy ahp and its applications. *European Journal of Operational Research*, 186(2):735–747, 2008.

[80] Amos Tversky and Itamar Simonson. Context-dependent preferences. *Management Science*, 39(10):1179–1189, 1993.

[81] Francisco Herrera and Luis Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.