ISSN: 0067-2904

Fuzzy Entropy in Adaptive Fuzzy Weighted Linear Regression Analysis with Application to Estimate Infant Mortality Rate

Ahmad Farouk Abbas*, Mohammed Jassim Mohammed
Department of Statistics, College of Administration and Economics, Baghdad University, Baghdad, Iraq.

Abstract
An adaptive fuzzy weighted linear regression model in which the output is based on the position and entropy of quadruple fuzzy numbers had dealt with. The solution of the adaptive models is established in terms of the iterative fuzzy least squares by introducing a new suitable metric which takes into account the types of the influence of different imprecisions. Furthermore, the applicability of the model is made by attempting to estimate the fuzzy infant mortality rate in Iraq using a selective set of inputs.

Keywords: Quadruple-FN, Fuzzy entropy, Fuzzy least squares, Non-negative least squares, Infant mortality rate.

الأنتروبي الضبابي في تحليل الإنحدار الخطي الضبابي الموزون المكيف مع تطبيق لتقدير معدل وفيات الرضع

أحمد فاروق عباس*، محمد جاسم محمد
قسم الأحصاء، كلية الإدارة والأقتصاد، جامعة بغداد، بغداد، العراق.

الخلاصة
تم التعامل في هذا البحث مع أنموذج خطي ضبابي موزون مكيف ذو مخرج يعتمد على الموقع والأنتروبي للأرقام الضبابية الرباعية إذ تم وضع الحل للأنموذج المكيف بالأعتماد على المربعات الصغرى الضبابية التكرارية عن طريق أستعمال معيار مناسب جديد تم تعريفهُ ليأخذ بنظر الأعتبار تأثير أنواع عدم الدقة المختلفة. علاوة على ذلك جرى تطبيق الأنموذج عبر محاولة لتقدير معدلات وفيات الرضع الضبابية في العراق بوساطة مجموعة من المدخلات المختارة.

1. Introduction

Regression analysis is an important mathematical tool used in wide range of sciences to investigate the relationship between a certain phenomenon (output) and a set of arbitrary inputs under the frame of uncertainty. The uncertainty of classical techniques in estimation and analysis of the unknown parameters attributed to the randomness of the relationship between output and inputs. Recently many studies began to take into account other types of uncertainty sources like vagueness, ambiguity, and imprecision whom affect statistical reasoning in regression models and perhaps not surprisingly to adopt the fuzzy theory to treat these types of uncertainty in what so called fuzzy regression models.

Based on the literature of fuzzy regression modeling, we can roughly categorize it into two kinds. One is the mathematical programming methods and the other one is the least squares methods.

*Email: areej_mussab@yahoo.com

Although the least squares methods are more complex computationally but it gives a better-estimated output compared to mathematical programming methods [1]. The first adaptive fuzzy linear regression model (AFLRM) was proposed by D'Urso and Gastaldi [2] and later extended by Coppi et.al [3] became one of the classical fuzzy linear regression models where the estimation procedure is the using of least squares methods. Usually, fuzzy numbers are used to present fuzzy data in fuzzy regression models where they are often determined by the core set, the left spread and the right spread values, as for the current study we used a different approach to determining fuzzy numbers by position and entropy (left entropy and right entropy). Based on the idea similar to that of AFLRM we construct the model by means of the position and entropy of quadruple fuzzy number with adding some weights to reduce the variation. The rest of the paper is organized as follows. In Section 2 we start with the definitions of various terms relating to fuzzy sets and numbers which they will be used later in the context of this paper. In Section 3 the adaptive fuzzy weighted linear regression model is introduced. Finally, we present some numerical application to estimate the fuzzy infant mortality rate in Iraq and make conclusions in Section 4 and 5 respectively.

2. Mathematical Preliminaries

2.1 Fuzzy Sets and Numbers

Definition (2.1.1) [4, 5]: Let $\mathcal{K}(\mathbb{R}^m) \neq \emptyset$ be the universe of discourse of fuzzy sets, then a fuzzy set $\mathcal{A}$ is a subset of the m-dimensional real space $\mathbb{R}^m$ characterized by the mapping $\mu_{\mathcal{A}}(x): \mathbb{R}^m \longrightarrow [0,1]$, so called the membership function which describes the membership degree of x in $\mathcal{A}$, and $\mathcal{A}$ is defined as a set of ordered pairs: $\mathcal{A} = \{(x, \mu_{\mathcal{A}}(x)) | x \in \mathbb{R}^m \longrightarrow [0,1]\}$.

In the following some basic properties of fuzzy set $\mathcal{A}$ which defined as crisp sets:

1. α −Level: $\mathcal{A}^\alpha = \{x \in \mathbb{R}^m | \mu_{\mathcal{A}}(x) \geq \alpha\}\ \forall\, \alpha \in [0,1]$
2. Support :$\mathcal{S}(\mathcal{A}) = \{x \in \mathbb{R}^m | \mu_{\mathcal{A}}(x) > 0\}$
3. Core (sometimes called as Kernel): $\mathcal{C}(\mathcal{A}) = \{x \in \mathbb{R}^m | \mu_{\mathcal{A}}(x) = 1\}$

Definition (2.1.2) [6]: Let $\mathcal{F}(\mathbb{R}^m) \neq \emptyset$ be the universe of discourse of fuzzy numbers then a fuzzy number $\mathcal{A} \in \mathcal{F}(\mathbb{R}^m)$ which identify with the membership function $\mu_{\mathcal{A}}(x): \mathbb{R}^m \longrightarrow [0,1]$, must satisfy the following:

1. Normal: $\exists x_o \ni \mu_{\mathcal{A}}(x_o) = \{x_o \in \mathbb{R}^m | \mu_{\mathcal{A}}(x_o) = 1\}$.
2. α −Level of fuzzy set $\mathcal{A}$ with a $\mu_{\mathcal{A}}(x)$ is a compact set: $\mu_{\mathcal{A}}(\alpha) = \{x \in \mathbb{R}^m | \mu_{\mathcal{A}}(x) \geq \alpha\}$ if $\alpha > 0$ and $\mu_{\mathcal{A}}(\alpha) = cl\{x \in \mathbb{R}^m | \mu_{\mathcal{A}}(x) > 0\}$ if $\alpha = 0\ \ \forall \alpha \in [0,1]$.
3. Upper semicontinuous: $\lim_{x \to \tau} sup(\mu_{\mathcal{A}}(x)) \leq \mu_{\mathcal{A}}(\tau),\ -\infty < \tau < \infty$.
4. Quasi-Convex: $\mu_{\mathcal{A}}(\rho x_1 + (1-\rho)x_2) \geq min(\mu_{\mathcal{A}}(x_1), \mu_{\mathcal{A}}(x_2))\ \forall \rho \in [0,1]$.

As a special case of fuzzy numbers, namely what called LR-fuzzy numbers (LR-FN's) which defined by two functions $L: \mathbb{R}^+ \to [0,1]$ and $R: \mathbb{R}^+ \to [0,1]$, where $L(x)$ and $R(x)$ are monotonic non increasing functions satisfies: $L(x) = R(x) = 0\ , if\ x \geq 1$ and $L(x) = R(x) = 1\ , if\ x \leq 0$ [5]. In general the membership function of a LR-FN can be represented as follows:

$$\mu_{\mathcal{A}}(x) \triangleq \begin{cases} L(x) & \text{if } x < m^l \\ 1 & \text{if } x \in [m^l, m^r] \\ R(x) & \text{if } m^r > x \end{cases} \quad (1)$$

where both of m^l, m^r represent the core set and $L(x), R(x) \in [0,1]$. Many special cases can be raised from LR-FN's above where one/both of the left or/and right functions are equal to zero $L(x), R(x) = 0$. Moreover, the left and right points of the core set could be equal $m^l = m^r$.

Definition (2.1.3) [4]: A fuzzy number $\mathcal{Y}^R = (\omega; \nu, \psi, \delta)_{LR} \in \mathcal{F}(\mathbb{R}^4)$ is defined as Quadruple-FN (QFN) with a R order of curved shape of membership function if its membership function is given by:

$$\mu_{\mathcal{Y}}{}^{R}(x) \triangleq \begin{cases} 1-\left(\frac{(\omega-\nu)-x}{\psi}\right)^{R} & \text{if } x \in [(\omega-\nu)-\psi,(\omega-\nu)) \\ 1 & \text{if } x \in [(\omega-\nu),(\omega+\nu)] \\ 1-\left(\frac{x-(\omega+\nu)}{\delta}\right)^{R} & \text{if } x \in ((\omega+\nu),(\omega+\nu)+\delta] \\ 0 & Otherwise \end{cases} \quad (2)$$

where ω represent mean value of core points, $\nu,\psi,\delta \in \mathbb{R}^{+}$ represent the spread values and $R \in (0,\infty)$. Obviously QFN is totally identified by its core set $\mathcal{C}(\mathcal{Y}^{R}) = \{(\omega-\nu),(\omega+\nu)\}$ and support set $\mathcal{S}(\mathcal{Y}^{R}) = \{(\omega-\nu)-\psi,(\omega+\nu)+\delta\}$, and for any $R \in (0,\infty)$ the left and right functions can be defined as: $L(x) = R(x) = 1-|x|^{R}, 0 \le x \le 1$. Choosing the right R order is really important in some applications especially the ones that utilize from the influence of imprecision in fuzzy numbers. In particular, if $R \in (0,1)$ then imprecision is decreasing aggressively and if $R = 1$ then imprecision is decreasing linearly while if $R \in (1,\infty)$ then imprecision decreasing slowly [7].

Proposition (2.1.1): Whether if $R \longrightarrow 0$ or $R \longrightarrow \infty$ the QFN still at its simplest an interval number.

Proof: assuming $\nu,\psi,\delta \neq 0$ the proof is explicit as follows

$$\lim_{R\to 0} \mu_{\mathcal{Y}}{}^{R}(x) \triangleq \begin{cases} 1 & \text{if } x \in [(\omega-\nu),(\omega+\nu)] \\ 0 & Otherwise \end{cases}$$

And

$$\lim_{R\to \infty} \mu_{\mathcal{Y}}{}^{R}(x) \triangleq \begin{cases} 1 & \text{if } x \in [(\omega-\nu)-\psi,(\omega-\nu),(\omega+\nu),(\omega+\nu)+\delta] \\ 0 & Otherwise \end{cases}$$ □

The above proposition (2.1.1) stated that if $R \longrightarrow 0$ the QFN will reduced to an interval over core set, while if $R \longrightarrow \infty$ the QFN will reduced to an interval over support set and this can be generalized to complex cases when the left $L(x)$ and right $R(x)$ functions of membership function have different orders of curved shapes.

Remark (2.1.1): If $\psi \neq \omega$ then the QFN is called Asymmetric QFN (LR-QFN), but when $\psi = \omega$ the QFN is called Symmetric QFN (LL-QFN) and denoted by $\mathcal{Y}^{R} = (\omega;\nu,\psi)_{LL}$, this can be represented geometrically as illustrated in Figure -1.

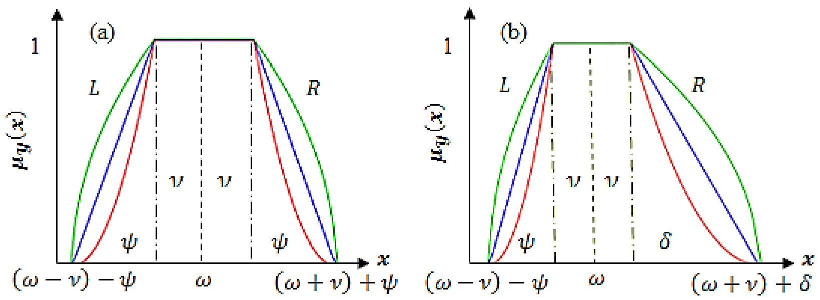


Figure 1- Geometric representation of **(a)** Symmetric QFN **(b)** Asymmetric QFN, with the different R order.

Although the fuzziness of fuzzy numbers described by the membership functions. But, Kao and Lin [8] proposed a different approach using simple indices for fuzziness instead of using membership functions. The idea was fair since the fuzzy numbers describes as a subsets of real line whose highest membership function clustered around given real points and the membership functions which represent the fuzziness are monotonic on both sides of these given points, they replaced these points by position of fuzzy number and the fuzziness interpreted by applying fuzzy entropy where fuzzy entropy functions are also monotonic on both sides of any fuzzy numbers.

Definition (2.1.4) [8, 9]: Fuzzy entropy is a measure of fuzziness represent the average of vagueness and ambiguity in making any decisions within the fuzzy environment, also it is defined as the amount of information that lost in the process of transformation from crisp set to fuzzy set. Any fuzzy entropy function should satisfy the following conditions:

1. Certainty: $E(\mathcal{Y}) = 0$, if $\mu_{\mathcal{Y}}(x) = 0 \; or \; 1$.
2. Maximality: $E(\mathcal{Y}) = 1$, if $\mu_{\mathcal{Y}}(x) = 0.5$.
3. Resolution: $E(\mathcal{Y}) \geq E(\mathcal{Z})$, if $\mu_{\mathcal{Z}}(x) \geq \mu_{\mathcal{Y}}(x) \geq 0.5$ and $\mu_{\mathcal{Z}}(x) \leq \mu_{\mathcal{Y}}(x) \leq 0.5$.
4. Symmetry: $E(\mathcal{Y}) = E(\mathcal{Y}^c) \Longrightarrow \; \mu_{\mathcal{Y}^c}(x) = 1 - \mu_{\mathcal{Y}}(x)$.

Let $e\left(\mu_{\mathcal{Y}}(x)\right): [0,1] \longrightarrow [0,1]$ be a fuzzy entropy of a fuzzy number $\mathcal{Y}$ that is at fixed x, the entropy e is monotonically increasing on the interval $[0, 1/2]$ and monotonically decreasing on the other interval $[1/2\,,1]$, the functions e could take several forms but the most fundamental ones are given below[10]:

$$e\left(\mu_{\mathcal{Y}}(x)\right) \triangleq \mu_{\mathcal{Y}}(x)\, ln\left({}^{1}/_{\mu_{\mathcal{Y}}(x)}\right) + (1 - \mu_{\mathcal{Y}}(x))\, ln\left({}^{1}/_{1 - \mu_{\mathcal{Y}}(x)}\right) \tag{3}$$

$$e\left(\mu_{\mathcal{Y}}(x)\right) \triangleq 4\mu_{\mathcal{Y}}(x) - 4\mu_{\mathcal{Y}}(x)^2 \tag{4}$$

where (3) defined as Shannon function and (4) represent the Logistic differential equation. To determine a global entropy whose independent of x values the above functions has to be integrated over the entire space of x (assuming the integration does make sense).This yields

$$E\left(\mathcal{Y}^R{}_{(x)}\right) \triangleq \int_{x\,\in\,\mathcal{X}(\mathbb{R}^m)} e\left(\mu_{\mathcal{A}_{(x)}}\right)\mathcal{P}(x)\,dx \tag{5}$$

where $\mathcal{P}(x)$ denotes a probability density function of the available data set defined over $(\mathbb{R}^m)$. Usually, $\mathcal{P}(x)$ assumed to be uniform distribution (constant over given interval i.e $\mathcal{P}(x) = k$). Depending on (5) fuzzy entropy of QFN will be partitioned into two components one for the increasing part $[(\omega - \nu) - \psi, (\omega - \nu))$ and the other for the decreasing part $((\omega + \nu), (\omega + \nu) + \delta]$. This yields

$$E\left(\mathcal{Y}^R{}_{(x)}\right) \triangleq k\begin{bmatrix} \displaystyle\int_{x\in[(\omega-\nu)-\psi,(\omega-\nu))} e\left(\mu_{\mathcal{Y}}{}^R(x)\right)\,dx \\ \displaystyle+ \int_{x\in((\omega+\nu),(\omega+\nu)+\delta]} e\left(\mu_{\mathcal{Y}}{}^R(x)\right)\,dx \end{bmatrix} \tag{6}$$

The first term in (6) is defined as left fuzzy entropy while the second term is defined as right fuzzy entropy. In a simple way we can rewrite (6) as $E\left(\mathcal{Y}^R{}_{(x)}\right) \triangleq e_{\mathcal{L}} + e_{\mathcal{R}}$ where $e_{\mathcal{L}} \triangleq \int_{-\infty}^{m} e\big(L(x)\big)\,dx < +\infty$ and $e_{\mathcal{R}} \triangleq \int_{m}^{\infty} e\big(R(x)\big)\,dx < +\infty$, here m denote the interval where $\mu_{\mathcal{Y}}{}^R(x) = 1$. Also It should be pointed out that the entropy function is completely depends on the spread of fuzzy number and if we use LL-QFN then the left and right entropy will be equal $e_{\mathcal{L}} = e_{\mathcal{R}}$.

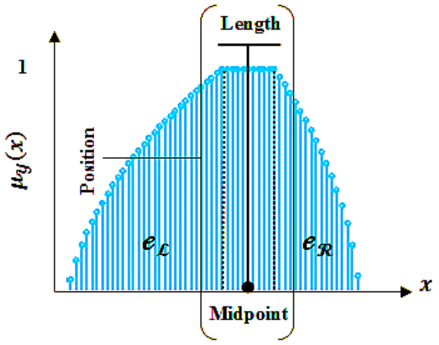


Figure 2- Geometric representation of QFN by means of the position and the entropies.

Ignoring the probability distribution $\mathcal{P}(x)$ the QFN's can be characterized uniquely by four crisp attributes (as shown in Figure-2): the position where ${\mu_{\mathcal{Y}}}^R(x) = 1$ whose represented by the length and the mean value of the interval $[(\omega - \nu), (\omega + \nu)]$ and by its fuzziness whose represented by the left entropy $e_{\mathcal{L}}$ and right entropy $e_{\mathcal{R}}$. That's it $\mathcal{Y}^R = \{\xi, \mathcal{s}, e_{\mathcal{L}}, e_{\mathcal{R}}\}$, where: $\xi = \text{mean} = ((\omega - \nu) + (\omega + \nu))/2 = \omega$ and $\mathcal{s} = \text{length} = (\omega + \nu) - (\omega - \nu) = 2\nu$ [8].

1.2 Scalar Metric

Definition (2.2.1): A positive map $\Delta: \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^+$, is called a scalar metric (in short just a metric) in $\mathbb{R}^m$ and $(\mathbb{R}^m, \Delta)$ is called a metric space, if Δ satisfies the following conditions for all $\mathcal{Y}, \mathcal{Y}^*, \mathcal{Y}' \in \mathbb{R}^m$:

1. Non-negativity (and Uniqueness): $\Delta(\mathcal{Y}, \mathcal{Y}^*) \geq 0$ (0 if $\mathcal{Y} = \mathcal{Y}^*$).

2. Symmetry: $\Delta(\mathcal{Y}, \mathcal{Y}^*) = \Delta(\mathcal{Y}^*, \mathcal{Y})$.

3. Triangular inequality: $\Delta(\mathcal{Y}, \mathcal{Y}^*) \leq \Delta(\mathcal{Y}, \mathcal{Y}') + \Delta(\mathcal{Y}', \mathcal{Y}^*)$.

Remark (2.2.1): If uniqueness condition did not satisfy then the metric is called a Pseudometric, while if triangular inequality did not satisfy then the metric is called a Semimetric.

Metrics are used as a measure of the degree of closeness between two points or in general between two sets whether it is crisp or fuzzy, driving appropriate metric for fuzzy sets consider the key for better performance in many application like regression, clustering, ranking etc. The metrics for fuzzy sets usually represent a generalization of the classical metrics, these kind of metrics are tolerable in data analysis. However, what worth to be mentioned that the problem of defining a satisfactory metric for any purpose in fuzzy data did not solved yet. In the literature, various metrics have been defined by different aspect, but the most popular of them are a generalization of two classical metrics between compact and convex set and this generalization made using α −Level which in turn represent a compact and convex sets [11]. First let start with the Hausdorff metric

$$\Delta_H(M,N) \triangleq \begin{cases} \sqrt[h]{\int_0^1 \left(\Delta_H(M^\alpha, N^\alpha)\right)^h d\alpha}, & h \in [1, \infty) \\ \sup\limits_{\alpha \in [0,1)} \Delta_H(M^\alpha, N^\alpha), & h = \infty \end{cases} \tag{7}$$

where M^α and N^α are the α −Level sets of fuzzy sets M and N respectively. The second metric is defined in L^2 space in terms of support function s as follows

$$\Delta_S \triangleq \sqrt{p \int_0^1 \int_{S^{p-1}} v\left(s_M(u,\alpha) - s_N(u,\alpha)\right)^2 du\, d\alpha} \tag{8}$$

where $s_M(u,\alpha) = \sup_{m^\alpha \in M^\alpha} \langle m^\alpha, u \rangle \ \forall \alpha \in [0,1];\ u \in S^{p-1}$, S^{p-1} is the $(p-1)$-dimensional unit sphere in $\mathbb{R}^p$and v is Lebesgue measure. Although previous metrics are defined in general for fuzzy sets but also can be simplified using specific kind of fuzzy number. However, the Hausdorff metric give to each points in fuzzy sets the same weight and there is no presence for any influence of imprecision type while the latter one solved this problem but defined in term of support function this make both metrics are inadequate in case of fuzzy numbers being characterized by position and entropy functions. Thus we introduced a new type of weighted metric defined in L^2 space between two fuzzy numbers, specifically between two QFN's which take into consideration the shape of the membership curve of QFN in being linear, quadratic, exponential or any other else, as well it is easy in computation.

Proposition (2.2.1): Given $\mathcal{Y}^R = \{\xi, s, e_{\mathcal{L}}, e_{\mathcal{R}}\}, \mathcal{Y}^{R*} = \{\xi^*, s^*, e_{\mathcal{L}}^*, e_{\mathcal{R}}^*\} \in \mathcal{F}(\mathbb{R}^4)$ as observed in a set of *n* units, a new weighted metric $\Delta: \mathcal{F}(\mathbb{R}^4) \times \mathcal{F}(\mathbb{R}^4) \to \mathbb{R}^+$ is defined as follows:

$$\Delta = \sqrt{w_\xi \|\xi - \xi^*\|^2 + w_s \|s - s^*\|^2 + w_{\mathcal{R}} \|(\Gamma e_{\mathcal{R}} - \Gamma e_{\mathcal{R}}^*)\|^2 + w_{\mathcal{L}} \|(\Pi e_{\mathcal{L}} + \Pi e_{\mathcal{L}}^*)\|^2} \tag{9}$$

where $w_\xi, w_s, w_{\mathcal{L}}, w_{\mathcal{R}} \in \mathbb{R}^+$ represent an arbitrary weights, $\Gamma = diag(\mathcal{J}_{\mathcal{R}})$ and $\Pi = diag(\mathcal{J}_{\mathcal{L}})$ are n by n positive definite matrices and $\mathcal{J}_{\mathcal{R}} = \int_0^1 R^{-1}(\varpi) d\varpi$ and $\mathcal{J}_{\mathcal{L}} = \int_0^1 L^{-1}(\varpi)\, d\varpi$ are denote the impact of the order of curve shape of the membership function on the metric, the type of imprecision depending on membership function. Specifically, the imprecision of QFN's is represent linearly, quadratic and square root when R=1,2 and 0.5 respectively.

Proof:It is easy to check conditions 1 – 2 in definition (2.2.1), now let $\mathcal{Y}^{R'} = \{\xi', s', e_{\mathcal{R}}', e_{\mathcal{L}}'\}$ be a 3rd QFN, for the triangular inequality we have

$$\Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) = w_\xi \|[\xi - \xi'] + [\xi' - \xi^*]\|^2 + w_s \|[s - s'] + [s' - s^*]\|^2$$
$$+ w_{\mathcal{R}} \|[\Gamma e_{\mathcal{R}} - \Gamma e_{\mathcal{R}}'] + [\Gamma e_{\mathcal{R}}' - \Gamma e_{\mathcal{R}}^*]\|^2 + w_{\mathcal{L}} \|[\Pi e_{\mathcal{L}} - \Pi e_{\mathcal{L}}'] + [\Pi e_{\mathcal{L}}' - \Pi e_{\mathcal{L}}^*]\|^2$$
$$\therefore \Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) = \Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R'}) + \Delta^2(w; \mathcal{Y}^{R'}, \mathcal{Y}^{R*}) + 2w_\xi(\xi - \xi')(\xi' - \xi^*)$$
$$+ 2w_s(s - s')(s' - s^*) + 2w_{\mathcal{R}}(\Gamma e_{\mathcal{R}} - \Gamma e_{\mathcal{R}}')(\Gamma e_{\mathcal{R}}' - \Gamma e_{\mathcal{R}}^*) + 2w_{\mathcal{L}}(\Pi e_{\mathcal{L}} - \Pi e_{\mathcal{L}}')(\Pi e_{\mathcal{L}}' - \Pi e_{\mathcal{L}}^*)$$

by Cauchy–Schwarz inequality

$$\Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) \le \Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R'}) + \Delta^2(w; \mathcal{Y}^{R'}, \mathcal{Y}^{R*}) + 2\Delta(w; \mathcal{Y}^R, \mathcal{Y}^{R'})\Delta(w; \mathcal{Y}^{R'}, \mathcal{Y}^{R*})$$
$$\Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) \le [\Delta(w; \mathcal{Y}^R, \mathcal{Y}^{R'}) + \Delta(w; \mathcal{Y}^{R'}, \mathcal{Y}^{R*})]^2$$

Thus

$$\Delta(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) \le \Delta(w; \mathcal{Y}^R, \mathcal{Y}^{R'}) + \Delta(w; \mathcal{Y}^{R'}, \mathcal{Y}^{R*})$$ □

Theorem (2.2.1): $(\mathcal{F}(\mathbb{R}^4), \Delta)$ is a complete metric space

Proof. Let $\{\mathcal{Y}_n^R\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}$ where $\mathcal{Y}_n^R = \{\xi_n, s_n, e_{\mathcal{L}n}, e_{\mathcal{R}n}\}$ and $w_\xi, w_s, w_{\mathcal{L}}, w_{\mathcal{R}} \in \mathbb{R}^+$(constants). Then all of $\{\xi_n\}_{n\in\mathbb{N}}$, $\{s_n\}_{n\in\mathbb{N}}$, $\{e_{\mathcal{L}n}\}_{n\in\mathbb{N}}$ and $\{e_{\mathcal{R}n}\}_{n\in\mathbb{N}}$ must be Cauchy sequences in $\mathbb{R}$, let $\epsilon > 0$ $\exists$ n_o $\ni$ $\Delta(\mathcal{Y}_n^R, \mathcal{Y}_m^R) < \epsilon$ $\forall$ $n, m > n_o \to \sqrt{w_\xi}|\xi_n - \xi_m| \le \Delta(\mathcal{Y}_n^R, \mathcal{Y}_m^R) < \epsilon$ $\therefore \{\xi_n\}_{n\in\mathbb{N}}$ is a Cauchy sequance in $\mathbb{R}$ and $\sqrt{w_s}|s_n - s_m| \le \Delta(\mathcal{Y}_n^R, \mathcal{Y}_m^R) < \epsilon \therefore \{s_n\}_{n\in\mathbb{N}}$ is a Cauchy sequance in $\mathbb{R}$.

Since Γ and Π are positive definite diagonal matrices, then similarly to above $\sqrt{w_{\mathcal{R}}}|\Gamma e_{\mathcal{R}n} - \Gamma e_{\mathcal{R}n}| \le \Delta(\mathcal{Y}_n^R, \mathcal{Y}_m^R) < \epsilon$ $\therefore \{e_{\mathcal{R}n}\}_{n\in\mathbb{N}}$ is a Cauchy sequance in $\mathbb{R}$ and

$\sqrt{w_{\mathcal{L}}}|\Pi e_{\mathcal{L}n} - \Pi e_{\mathcal{L}n}| \leq \Delta(\mathcal{Y}_n^R, \mathcal{Y}_m^R) < \epsilon \therefore \{e_{\mathcal{L}n}\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. Now to prove $\{\mathcal{Y}_n^R\}_{n\in\mathbb{N}}$ is convergent in $\mathcal{F}(\mathbb{R}^4)$ (i.e $\lim_{n\to\infty} \mathcal{Y}_n^R = \{\xi, s, e_{\mathcal{L}}, e_{\mathcal{R}}\}$), let $\epsilon > 0$ $\exists\, n_1 \ni$ $\sqrt{w_\xi}|\xi_n - \xi_m| \leq \epsilon, \exists\, n_2 \ni$ $\sqrt{w_s}|s_n - s_m| \leq \epsilon,$ $\exists\, n_3 \ni$ $\sqrt{w_{\mathcal{R}}}|\Gamma e_{\mathcal{R}n} - \Gamma e_{\mathcal{R}n}| \leq \epsilon$ and $\exists\, n_4 \ni$ $\sqrt{w_{\mathcal{L}}}|\Pi e_{\mathcal{L}n} - \Pi e_{\mathcal{L}n}| \leq \epsilon,$ $\forall$ $n > n_1, n_2, n_3, n_4$. Set $n_o = max\{n_1, n_2, n_3, n_4\}$ $\therefore \exists$ $n > n_o$ $\ni$ $\Delta^2(\mathcal{Y}_n^R, \mathcal{Y}^R) = \sqrt{w_\xi}|\xi_n - \xi_m|^2 + w_s|s_n - s_m|^2$ $+ w_{\mathcal{R}}|\Gamma e_{\mathcal{R}n} - \Gamma e_{\mathcal{R}n}|^2 + w_{\mathcal{L}}|\Pi e_{\mathcal{L}n} - \Pi e_{\mathcal{L}n}|^2 < \epsilon^2$. Thus $\lim_{n\to\infty} \mathcal{Y}_n^R = \mathcal{Y}^R$ □

Remark (2.2.2): The reason behind adding the arbitrary weights was to reduce the dominance of measurements difference and lessen the variations between the points in QFN data in calculating the metric for providing more accuracy. This would be useful especially if we use the metric in economical or biological modeling. Notice if we set $w_\xi = w_s = w_{\mathcal{L}} = w_{\mathcal{R}} = 1$ then we cancel any effect of the weights while in case of LL-TFN the weights of left and right entropies must be equal $w_{\mathcal{L}} = w_{\mathcal{R}}$.

Remark (2.2.3): If the spreads of QFN's equal to zero i.e. $\mathcal{Y}^R = (\xi; 0,0,0)_{LR}$ and $\mathcal{Y}^{R*} = (\xi^*; 0,0,0)_{LR}$, then the QFN's become crisp numbers and the metric (9) reduced to the Manhattan distance (setting $w_\xi = w_s = w_{\mathcal{L}} = w_{\mathcal{R}} = 1$ without loss of generality) $\Delta(\mathcal{Y}^R, \mathcal{Y}^{R*}) = |\xi - \xi^*|$, which indicates that the new metric is a generalization of the Euclidean distance.

Remark (2.2.4): There exist some points when the left and/or right curve shape of the membership function of $\mathcal{Y}^R, \mathcal{Y}^{R*}$ are different but $\Delta(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) = 0$, this makes Δ a Pseudometric since $\mathcal{Y}^R \neq \mathcal{Y}^{R*}$, so a prerequisite to the metric (9) is that $\mathcal{Y}^R$ and $\mathcal{Y}^{R*}$ should have exactly the same curve shape.

3. Adaptive Fuzzy Linear Regression Model (AFLRM)

Let the input variables $f_{i1}, f_{i2}, \dots, f_{ik}$ be crisp and the output variable $\mathcal{Y}_i^R = (\xi_i; \nu_i, \psi_i, \delta_i)_{LR}$ be QFN's observed on a sample of size n (i.e. $i = 1,2, \dots, n$), the usual (non adaptive FLR) method to estimate the fuzzy regression line is to construct four lines for each of the core set and the two spreads:

$$\mathcal{Y} = \theta_o \oplus (\theta_1 f_1) \oplus (\theta_2 f_2) \oplus \dots \oplus (\theta_k f_k) \qquad (10)$$

The above model consist that there is no influence of relation between the four models, under some constraints of fuzzy arithmetic operations each line can be estimated by finding the estimation of ordinary least squares method, D'Urso and Glatdsui [3] proposed another type of fuzzy regression model which they considered that the dynamic of left and right spreads depends in somehow on the estimated core, this model is so-called adaptive fuzzy linear regression model (AFLRM), later this were developed by D'Urso [12] and Coppi et.al [4], the method can be summarized by building four linear models two for the core set $\mathcal{C}(\mathcal{Y}^R) = \{C_1 = (\omega - \nu), C_2 = (\omega + \nu)\}$ by means of a classical regression models and simultaneously modeling the two others for spreads by the linear relation between spreads and core set, in other words build the spreads as a regression models of core set as follows:

$$C_1^* = \mathcal{F}_x \Omega_1$$
$$C_2^* = \mathcal{F}_x \Omega_2$$
$$\psi^* = C_1^* \lambda_1 + C_2^* \lambda_2 + \mathbf{1}\lambda_o$$
$$\delta^* = C_1^* \alpha_1 + C_2^* \alpha_2 + \mathbf{1}\alpha_o \qquad (11)$$

where $C_1^*, C_2^*, \psi^*, \delta^*$ are n component vectors represent the estimated outputs of the spreads and core set, $\mathcal{F}_x$ is the design matrix, Ω_1, Ω_2 are p component vectors of core set parameters, $(\delta_1, \delta_2, \delta_o, \alpha_1, \alpha_2, \alpha_o)$ are scalars represent the parameters between the observation of spreads and the estimated core and $\mathbf{1}$ is a n component vector of ones. In the same manner we use the AFLRM but instead we used QFN as described by position and entropy. Thus (11) can be parameterized as linear regression models by using the outputs as $\mathcal{Y}_i^R = \{\xi_i, s_i, e_{\mathcal{L}i}, e_{\mathcal{R}i}\}$. as follows:

$$\xi^* = \mathcal{F}_x\theta_1$$
$$s^* = \mathcal{F}_x\theta_2$$
$$e_{\mathcal{L}}^* = \xi^*\tau_1 + s^*\tau_2 + \mathbf{1}\tau_o$$
$$e_{\mathcal{R}}^* = \xi^*\varphi_1 + s^*\varphi_2 + \mathbf{1}\varphi_o \qquad (12)$$

where $\xi^*, s^*, e_{\mathcal{L}}^*, e_{\mathcal{R}}^*$ are n component vectors represent the estimated outputs of $\mathcal{Y}_i^R = \{\xi_i, s_i, e_{\mathcal{L}i}, e_{\mathcal{R}i}\}$, θ_1, θ_2 are p component vectors of position parameters and $(\tau_1, \tau_2, \tau_o, \varphi_1, \varphi_2, \varphi_o)$ are scalars represent the parameters between the observation of entropy and the estimated position. Despite this model is capable to incorporate the effect of the core on the spreads or in our case the effect of the position on the entropy but what worth mentioning as an advantage, is this AFLRM have smaller variance compared to the non adaptive fuzzy linear regression.

To estimate this model which consist of four sub-models the fuzzy least squares approach were applied. The objective of the fuzzy least squares estimation method is to estimate the parameters of the fuzzy regression model with least difference by minimizing the squared metric Δ^2 between observations $\mathcal{Y}_i^R = \{\xi_i, s_i, e_{\mathcal{L}i}, e_{\mathcal{R}i}\}$ and interpolated observations $\mathcal{Y}_i^{R*} = \{\xi_i^*, s_i^*, e_{\mathcal{L}i}^*, e_{\mathcal{R}i}^*\}$.

Theorem (3.1): The problem of $\min_{(\theta_1,\theta_2,\tau_1,\tau_2,\tau_o,\varphi_1,\varphi_2,\varphi_o)} \Delta^2(w; \theta_1, \theta_2, \tau_1, \tau_2, \tau_o, \varphi_1, \varphi_2, \varphi_o)$ admits a relative minimum which can be improved using an iterative algorithm.

Proof: Using the matrix form of the squared metric in (9)

$$\Delta^2(w; \mathcal{Y}^R, \mathcal{Y}^{R*}) = w_\xi(\xi - \xi^*)'(\xi - \xi^*) + w_s(s - s^*)'(s - s^*)$$
$$+ w_{\mathcal{R}}(\Gamma e_{\mathcal{R}} - \Gamma e_{\mathcal{R}}^*)'(\Gamma e_{\mathcal{R}} - \Gamma e_{\mathcal{R}}^*) + w_{\mathcal{L}}(\Pi e_{\mathcal{L}} + \Pi e_{\mathcal{L}}^*)'(\Pi e_{\mathcal{L}} + \Pi e_{\mathcal{L}}^*) \qquad (13)$$

Substituting the equations (12) in (13) and addressing the minimum problem we get:

$$\min_{(\theta_1,\theta_2,\tau_1,\tau_2,\tau_o,\varphi_1,\varphi_2,\varphi_o)} \Delta^2 = w_\xi(\xi'\xi - 2\xi'\mathcal{F}_x\theta_1 + \theta_1'\mathcal{F}_x'\mathcal{F}_x\theta_1) + w_s(s's - 2s'\mathcal{F}_x\theta_2 + \theta_2'\mathcal{F}_x'\mathcal{F}_x\theta_2)$$
$$+ w_{\mathcal{L}}\begin{pmatrix} e_{\mathcal{L}}'\Pi^2 e_{\mathcal{L}} - 2e_{\mathcal{L}}'\Pi^2(\mathcal{F}_x\theta_1\tau_1 + \mathcal{F}_x\theta_2\tau_2 + \mathbf{1}\tau_o) \\ +\tau_1^2\theta_1'\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_1 + \tau_2^2\theta_2'\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2 + \tau_o^2\mathbf{1}'\Pi^2\mathbf{1} \\ +2\tau_1\tau_2\theta_1'\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2 + 2\tau_1\theta_1'\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o + 2\tau_2\theta_2'\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o \end{pmatrix}$$
$$+ w_{\mathcal{R}}\begin{pmatrix} e_{\mathcal{R}}'\Gamma^2 e_{\mathcal{R}} - 2e_{\mathcal{R}}'\Gamma^2(\mathcal{F}_x\theta_1\varphi_1 + \mathcal{F}_x\theta_2\varphi_2 + \mathbf{1}\varphi_o) \\ +\varphi_1^2\theta_1'\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_1 + \varphi_2^2\theta_2'\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_2 + \varphi_o^2\mathbf{1}'\Gamma^2\mathbf{1} \\ +2\varphi_1\varphi_2\theta_1'\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_2 + 2\varphi_1\theta_1'\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o + 2\varphi_2\theta_2'\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o \end{pmatrix} \qquad (14)$$

In order to minimize Δ^2, the parameter estimates are obtained from the partial derivatives with respect to each $\theta_1, \theta_2, \tau_1, \tau_2, \varphi_1, \varphi_2, \eta_1, \eta_2$ associated with (14) being set equal to zero.

$$\frac{\partial\Delta^2}{\partial\theta_1} = w_\xi[-\xi'\mathcal{F}_x + \mathcal{F}_x'\mathcal{F}_x\theta_1] + w_{\mathcal{L}}\begin{bmatrix} -e_{\mathcal{L}}'\Pi^2\mathcal{F}_x\tau_2 + \tau_2^2\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_1 + \tau_2\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o \\ +\tau_1\tau_2\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2 + \tau_2\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o \end{bmatrix}$$
$$+ w_{\mathcal{R}}[-e_{\mathcal{R}}'\Gamma^2\mathcal{F}_x\varphi_2 + \varphi_2^2\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_1 + \varphi_2\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o + \varphi_1\varphi_2\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_2 + \varphi_2\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o] = 0_{p\times1}$$

$$\frac{\partial\Delta^2}{\partial\theta_2} = w_s[-s'\mathcal{F}_x + \mathcal{F}_x'\mathcal{F}_x\theta_2] + w_{\mathcal{L}}\begin{bmatrix} -e_{\mathcal{L}}'\Pi^2\mathcal{F}_x\tau_2 + \tau_2^2\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2 + \tau_2\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o \\ +\tau_1\tau_2\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_1 + \tau_2\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o \end{bmatrix}$$
$$+ w_{\mathcal{R}}[-e_{\mathcal{R}}'\Gamma^2\mathcal{F}_x\varphi_2 + \varphi_2^2\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_2 + \varphi_2\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o + \varphi_1\varphi_2\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_1 + \varphi_2\mathcal{F}_x'\Gamma^2\mathbf{1}\varphi_o] = 0_{p\times1}$$

$$\frac{\partial\Delta^2}{\partial\tau_1} = w_{\mathcal{L}}[-e_{\mathcal{L}}'\Pi^2\mathcal{F}_x\theta_1 + \tau_1\theta_1'\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_1 + \theta_1'\mathcal{F}_x'\Pi^2\mathbf{1}\tau_o + \tau_2\theta_1'\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2] = 0$$

$$\frac{\partial \Delta^2}{\partial \tau_2} = w_{\mathcal{L}}[-e_{\mathcal{L}}'\Pi^2 \mathcal{F}_x\theta_2 + \tau_2\theta_2'\mathcal{F}_x'\Pi^2 \mathcal{F}_x\theta_2 + \theta_2'\mathcal{F}_x'\Pi^2 \mathbf{1}\tau_o + \tau_1\theta_2'\mathcal{F}_x'\Pi^2 \mathcal{F}_x\theta_1] = 0$$
$$\frac{\partial \Delta^2}{\partial \varphi_1} = w_{\mathcal{R}}[-e_{\mathcal{R}}'\Gamma^2 \mathcal{F}_x\theta_1 + \varphi_1\theta_1'\mathcal{F}_x'\Gamma^2 \mathcal{F}_x\theta_1 + \theta_1'\mathcal{F}_x'\Gamma^2 \mathbf{1}\varphi_o + \varphi_2\theta_1'\mathcal{F}_x'\Gamma^2 \mathcal{F}_x\theta_2] = 0$$
$$\frac{\partial \Delta^2}{\partial \varphi_2} = w_{\mathcal{R}}[-e_{\mathcal{R}}'\Gamma^2 \mathcal{F}_x\theta_2 + \varphi_2\theta_2'\mathcal{F}_x'\Gamma^2 \mathcal{F}_x\theta_2 + \theta_2'\mathcal{F}_x'\Gamma^2 \mathbf{1}\varphi_o + \varphi_1\theta_2'\mathcal{F}_x'\Gamma^2 \mathcal{F}_x\theta_1] = 0$$
$$\frac{\partial \Delta^2}{\partial \tau_o} = w_{\mathcal{L}}[-e_{\mathcal{L}}'\Pi^2 \mathbf{1} + \tau_o\mathbf{1}'\Pi^2 \mathbf{1} + \tau_1\theta_1'\mathcal{F}_x'\Pi^2\mathbf{1} + \tau_2\theta_2'\mathcal{F}_x'\Pi^2\mathbf{1}] = 0$$
$$\frac{\partial \Delta^2}{\partial \varphi_o} = w_{\mathcal{R}}[-e_{\mathcal{R}}'\Gamma^2 \mathbf{1} + \varphi_o\mathbf{1}'\Gamma^2 \mathbf{1} + \varphi_1\theta_1'\mathcal{F}_x'\Gamma^2\mathbf{1} + \varphi_2\theta_2'\mathcal{F}_x'\Gamma^2\mathbf{1}] = 0$$

Thus, an iterative general solution of the above system can be represented by the following set of equations:

$$\theta_1^{(k+1)} = \begin{bmatrix} w_{\xi}\mathcal{F}_x'\mathcal{F}_x \\ +w_{\mathcal{L}}\left(\tau_1^{(k)}\right)^2 \mathcal{F}_x'\Pi^2\mathcal{F}_x \\ +w_{\mathcal{R}}\left(\varphi_1^{(k)}\right)^2 \mathcal{F}_x'\Gamma^2\mathcal{F}_x \end{bmatrix}^{-1} \begin{pmatrix} w_{\xi}\mathcal{F}_x'\xi + w_{\mathcal{L}}\tau_1^{(k)}\mathcal{F}_x'\Pi^2\left[e_{\mathcal{L}} - \tau_2^{(k)}\mathcal{F}_x\theta_2^{(k)} - \tau_o^{(k)}\mathbf{1}\right] \\ +w_{\mathcal{R}}\varphi_1^{(k)}\mathcal{F}_x'\Gamma^2\left[e_{\mathcal{R}} - \varphi_2^{(k)}\mathcal{F}_x\theta_2^{(k)} - \varphi_o^{(k)}\mathbf{1}\right] \end{pmatrix}$$

$$\theta_2^{(k+1)} = \begin{bmatrix} w_{s}\mathcal{F}_x'\mathcal{F}_x \\ +w_{\mathcal{L}}\left(\tau_2^{(k)}\right)^2 \mathcal{F}_x'\Pi^2\mathcal{F}_x \\ +w_{\mathcal{R}}\left(\varphi_2^{(k)}\right)^2 \mathcal{F}_x'\Gamma^2\mathcal{F}_x \end{bmatrix}^{-1} \begin{pmatrix} w_{s}\mathcal{F}_x's + w_{\mathcal{L}}\tau_2^{(k)}\mathcal{F}_x'\Pi^2\left[e_{\mathcal{L}} - \tau_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \tau_o^{(k)}\mathbf{1}\right] \\ +w_{\mathcal{R}}\varphi_2^{(k)}\mathcal{F}_x'\Gamma^2\left[e_{\mathcal{R}} - \varphi_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \varphi_o^{(k)}\mathbf{1}\right] \end{pmatrix}$$

$$\tau_1^{(k+1)} = \left(\theta_1^{(k)'}\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_1^{(k)}\right)^{-1}\theta_1^{(k)'}\mathcal{F}_x'\Pi^2\left(e_{\mathcal{L}} - \tau_2^{(k)}\mathcal{F}_x\theta_2^{(k)} - \tau_o^{(k)}\mathbf{1}\right)$$
$$\tau_2^{(k+1)} = \left(\theta_2^{(k)'}\mathcal{F}_x'\Pi^2\mathcal{F}_x\theta_2^{(k)}\right)^{-1}\theta_2^{(k)'}\mathcal{F}_x'\Pi^2\left(e_{\mathcal{L}} - \tau_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \tau_o^{(k)}\mathbf{1}\right)$$
$$\varphi_1^{(k+1)} = \left(\theta_1^{(k)'}\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_1^{(k)}\right)^{-1}\theta_1^{(k)'}\mathcal{F}_x'\Gamma^2\left(e_{\mathcal{R}} - \varphi_2^{(k)}\mathcal{F}_x\theta_2^{(k)} - \varphi_o^{(k)}\mathbf{1}\right)$$
$$\varphi_2^{(k+1)} = \left(\theta_2^{(k)'}\mathcal{F}_x'\Gamma^2\mathcal{F}_x\theta_2^{(k)}\right)^{-1}\theta_2^{(k)'}\mathcal{F}_x'\Gamma^2\left(e_{\mathcal{R}} - \varphi_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \varphi_o^{(k)}\mathbf{1}\right)$$
$$\tau_o^{(k+1)} = tr(diag(\mathcal{J}_{\mathcal{L}}^{-2}))\mathbf{1}'\Pi^2\left[e_{\mathcal{L}} - \tau_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \tau_2^{(k)}\mathcal{F}_x\theta_2^{(k)}\right]$$
$$\varphi_o^{(k+1)} = tr(diag(\mathcal{J}_{\mathcal{R}}^{-2}))\mathbf{1}'\Gamma^2\left[e_{\mathcal{R}} - \varphi_1^{(k)}\mathcal{F}_x\theta_1^{(k)} - \varphi_2^{(k)}\mathcal{F}_x\theta_2^{(k)}\right] \quad (15)$$

To solve the iterative equations (15) many algorithms can be adopted. In this paper, we used a simple one at certain threshold $\epsilon = 10^{-6}$, with two stages one for estimation and the other to test the stabilization of the local minimum optimization problem (the estimates). The algorithm can be illustrated as shown below:

Simple algorithm of iterative solution for AFLRM

Inputs: $\mathcal{F}_x \in \mathbb{R}^{p\times n},\ \xi \in \mathbb{R}^n,\ s, e_{\mathcal{L}}, e_{\mathcal{R}} \in (\mathbb{R}^+)^n$

Outputs: $\theta_1, \theta_2, \tau_1, \tau_2, \tau_o, \varphi_1, \varphi_2, \varphi_o \triangleq arg\ inf\left(\Delta^2(w; \theta_1, \theta_2, \tau_1, \tau_2, \tau_o, \varphi_1, \varphi_2, \varphi_o)\right)$

Initialization: $p^k = \{\theta_1{}^k, \theta_2{}^k, \tau_1{}^k, \tau_2{}^k, \tau_o{}^k, \varphi_1{}^k, \varphi_2{}^k, \varphi_o{}^k\} \neq \emptyset;\ p^k \in \mathbb{R};\ 1 \le k \le N$

Iterations:

- **First stage (Estimation procedure)**

→ First step: Use p^k to estimate p^{k+1}

→ Second step: Update p^k to p^{k+1}

→ Third step: $c = p^{k+1} - p^k$
→ Fourth step: a. If $c \leq \epsilon$ stop the loop, the estimates are p^{k+1}
b. If $c \nleq \epsilon$ return to second step and proceed until convergence is done in (a)

- **Second stage (Checking stabilization)**

→ First step: Choose different sets of p^k
→ Second step: Repeat the first stage until convergence criteria are met
→ Fourth step: If the estimates of the first stage equals the estimates of the second stage then the solution is stable else wise replace p^k until first and second stages are achieved

Obviously, the adaptive model (12) provides crisp parameters. But, since the outputs are fuzzy while the inputs are crisp, the fuzzy linear regression model involves an implicit fuzzy regression model in terms of fuzzy regression parameters. Thus, the crisp parameters of the adaptive model (12) it might involve a certain degree of impreciseness. In particular, we can extend the assessment of imprecision in [3] to our proposed model (12) to evaluate the imprecision due to the crisp parameters of the adaptive model. Following a similar procedure to the case with $\mathcal{Y}^R = \{\xi, s, e_{\mathcal{L}}, e_{\mathcal{R}}\}$ output and crisp inputs we can express the implicit model in the following way:

$$\xi = \mathcal{F}_x \beta_\xi$$
$$s = \mathcal{F}_x \beta_s$$
$$e_{\mathcal{L}} = |\mathcal{F}_x| \beta_{\mathcal{L}}$$
$$e_{\mathcal{R}} = |\mathcal{F}_x| \beta_{\mathcal{R}} \qquad (16)$$

where $|\mathcal{F}_x|$ is represent the design matrix of the absolute crisp inputs $|f_{ik}|$. The system of equations (16) may not verify the iterative solutions in (15), but can be utilized in order to get estimates of $\beta_\xi, \beta_s, \beta_{\mathcal{L}}, \beta_{\mathcal{R}}$ which are compatible with the parameters $\theta_1, \theta_2, \tau_1, \tau_2, \tau_o, \varphi_1, \varphi_2, \varphi_o$ in (12). That is

$$\xi^* = \mathcal{F}_x \beta_\xi + \varepsilon_\xi$$
$$s^* = \mathcal{F}_x \beta_s + \varepsilon_s$$
$$e_{\mathcal{L}}^* = |\mathcal{F}_x| \beta_{\mathcal{L}} + \varepsilon_{\mathcal{L}}$$
$$e_{\mathcal{R}}^* = |\mathcal{F}_x| \beta_{\mathcal{R}} + \varepsilon_{\mathcal{L}} \qquad (17)$$

where $\varepsilon_\xi, \varepsilon_s, \varepsilon_{\mathcal{L}}, \varepsilon_{\mathcal{L}}$ denote the vectors of residuals. Employing ordinary Least Squares, we can obtain compatible estimate of (12). For instance, the OLS estimates of β_ξ and β_s are:

$$\beta_\xi = (\mathcal{F}_x' \mathcal{F}_x)^{-1} \mathcal{F}_x' \xi^* = (\mathcal{F}_x' \mathcal{F}_x)^{-1} \mathcal{F}_x' \mathcal{F}_x \theta_1 = \theta_1$$
$$\beta_s = (\mathcal{F}_x' \mathcal{F}_x)^{-1} \mathcal{F}_x' s^* = (\mathcal{F}_x' \mathcal{F}_x)^{-1} \mathcal{F}_x' \mathcal{F}_x \theta_2 = \theta_2 \qquad (18)$$

This indicates that β_ξ and β_s match the estimates of iterative solution of θ_1 and θ_2, as for the estimates of the parameters of entropy sub-models $\beta_{\mathcal{L}}$ and $\beta_{\mathcal{R}}$ in (17) we need a positive estimates so that the multiplying of a positive matrix with a positive vector reduced to a positive vector (because $e_{\mathcal{L}}, e_{\mathcal{R}} \in (\mathbb{R}^+)^n$), one can adopt a constrained least squares problem due to avoid negative estimates. Namely, the non-negative Least Squares (NNLS) algorithm [13] under the Karush-Kuhn-Tucker (KKT) conditions to find an optimal solution in nonlinear (quadratic) programming. Given a design matrix $\mathcal{F}_x \in \mathbb{R}^{p \times n}$ and output $\gamma \in \mathbb{R}^n$, the problem to find a nonnegative vector $\beta \in (\mathbb{R}^+)^p$ to minimize the squared norm between observations γ and estimated observation $\mathcal{F}_x \beta$ can be represented as follows:

$$\min_\beta \Delta^2(\beta) = \frac{1}{2} ||\mathcal{F}_x \beta - \gamma||^2$$
$$subject\ to\ \beta \geq 0 \qquad (19)$$

Lawson and Hanson give the standard algorithm for the non-negative Least Squares (NNLS) which is an active set method whose based on only a small subset of active constraints i.e. satisfied the solution exactly. Assuming there are n inequality constraints in NNLS optimization problem. If the ith estimated parameter is negative, then the ith constraint is said to be active. Otherwise, the constraint is passive. The NNLS algorithm is given below, where $\mathcal{F}_x^{\,P}$ is a matrix associated with only the inputs currently in the passive set P.

Lawson and Hanson's algorithm for NNLS

Inputs: $\mathcal{F}_x \in \mathbb{R}^{p\times n},\ \gamma \in \mathbb{R}^n$

Outputs: $\beta \geq 0_p$ where $\beta \triangleq arg\ min\ \|\mathcal{F}_x\beta - \gamma\|^2$

Initialization: $P \neq \emptyset,\ \ Q = \{1,2,\dots,n\},\ \ \beta = 0_p,\ \ \mathrm{w} = \mathcal{F}_x{}'(\gamma - \mathcal{F}_x\beta)$

Iterations:

→ First step: Progress if $Q \neq \emptyset \wedge [max_{i\in Q}(w_i) > \epsilon]$

→ Second step: $j \triangleq arg\ max_{i\in Q}(w_i)$

→ Third step: Include the index j in P and remove it from Q

→ Fourth step: $s^P = \left(\mathcal{F}_x{}^{P\prime}\mathcal{F}_x{}^{P}\right)^{-1}\mathcal{F}_x{}^{P\prime}\gamma$

- Continue if $min\ (s^P) \leq 0$
- $\sigma = -min_{i\in P}(b_i/b_i - s_i)$
- $\beta \triangleq \beta + \sigma(s - \beta)$
- Update Q and P
- $s^P = \left(\mathcal{F}_x{}^{P\prime}\mathcal{F}_x{}^{P}\right)^{-1}\mathcal{F}_x{}^{P\prime}\gamma$
- $s^Q = 0$

→ Fifth step: $\beta = s$

→ Sixth step: $\mathrm{w} = \mathcal{F}_x{}'(\gamma - \mathcal{F}_x\beta)$

The performance of the AFLRM's in (12) has been evaluated using four goodness of fit indices to tell how well the fit of the model was. First, $\mathcal{R}^2$ to measure the proportion of the variation of output variable explained by the inputs and $\mathcal{R}^2_{adjusted}$ is also the same but based on the sample size and the number of inputs, where both of them defined as $\mathcal{R}^2 = \Delta^2(w; \mathcal{Y}^{R*}, \bar{\mathcal{Y}}^R)/\Delta^2(w; \mathcal{Y}^R, \bar{\mathcal{Y}}^R)$ and $\mathcal{R}^2_{adjusted} = 1 - (1 - \mathcal{R}^2)((n-1)/(n - \ddot{p}))$respectively, where: $\ddot{p}$ is the number of estimated parameters. Second, the average level set difference $\mathcal{ALSD}$ index is used to measure the mean difference between output and estimated output. Moreover, we proposed a new measure to obtain the reliability of prediction as follows:

$$\mathcal{AR} = n^{-1}\sum_{i=1}^{n} e^{-\Delta_i'} \tag{20}$$

where:

$$\Delta_i' = \frac{1}{n}\left(w_\xi|\xi_i - \xi_i{}^*| + w_s|s_i - s_i{}^*| + w_{\mathcal{R}}|\Gamma e_{\mathcal{R}i} - \Gamma e_{\mathcal{R}i}{}^*| + w_{\mathcal{L}}|\Pi e_{\mathcal{L}i} + \Pi e_{\mathcal{L}i}{}^*|\right)$$

Since $\Delta_i' \in [0,\infty)$ then $\mathcal{AR} \in [0,1]$, the closer $\mathcal{AR}$ values get to one the more high accuracy in model's prediction. A graphical analysis is also provided using Taylor diagram [14] which is used for summarizing how a set of patterns is close to match the reference pattern (observations) this quantified between any test pattern and the reference pattern by their correlation, their centered root mean squared difference and their standard deviations. The diagram can be visualized as a series of points on a polar coordinate system where the azimuth angle ϕ, which related to each point is such that $\cos(\phi)$ is equal to the correlation coefficient between the test pattern and reference pattern while the radius from the origin in the Taylor diagram represents the standard deviation of the test pattern, and the correlation between test pattern and reference pattern is given by the azimuthal position. The best test pattern to match the reference perfectly is the one whom had a radius equal to one and azimuth angle equal to zero. Taylor diagram can be described using the cosine law as follows:

$$\mathrm{E}'^2 = \sigma_f{}^2 + \sigma_t{}^2 - 2\sigma_f\sigma_t\cos(\phi) \tag{21}$$

where σ_f and σ_t represent the standard devation of reference and test patterns respectively; $\cos(\phi)$ represents the correlation coefficient between the test and reference patterns; E' is the centered root mean squared difference, E' is given for a reference f and a test t each one with n unit as below:

$$E' = \sqrt{n^{-1}\sum_{i=1}^{n}\left[(f_i - \bar{f}) - (t_i - \bar{t})\right]^2}$$

4. Practical Application: Infant Mortality Rate

The investigation for estimating infant mortality rate (IMR) models is not a new problem because of the interest in many disciplines like healthcare, demography and economics to name a few. In general mortality rate considered as one of the important indicators since most health measures that refer to incidence and/or prevalence of diseases are not available especially in countries with weak database they used to compare the economic and social development between countries. The high rate of mortality indicates a presence of danger to the population.

Attention was increased about IMR and even became the fourth-millennium development goals of the UN in an attempt to reduce it. A population with unhealthy infants harms the next generation in too many levels and it is natural to think that any factors that affect human development also affect IMR in the same country, and vice versa [15]. Infant mortality rate is mathematically defined as the number of infants dying before reaching one year of age, per 1,000 live births in a given year as follows [16]:

$$M_{<1}(t) = \left(\mathcal{D}_{<1}(t)/\bar{\mathcal{N}}_{<1}(t)\right) \times 1000 \qquad (22)$$

where $\mathcal{D}_{<1}(t)$ represents number of deaths of children before reaching one year of age at year t; $\bar{\mathcal{N}}_{<1}(t)$ represents risk exposure which estimated as an average population size of children before reaching one year of age in the middle of year t. Infant mortality rate models take many forms depending on the researchers, in this paper we chose to link regression mortality rate model that take into account the imprecision of the data with a set of relevant inputs. This can be represented by:

$$M_{<1}(t) \sim FPI + WS + GDP_{PPP} + TFR \qquad (23)$$

where FPI represent food production index which covers all the edible food crops that food crops that contain nutrients; WS represent the percentage of the population with access to improved water source; TFR represent the total fertility rate (number of births per woman); GDP_{PPP} represent the. gross domestic product converted to international dollars using purchasing power parity (PPP) rates method. The data of fuzzy IMR and the crisp set of inputs are given in Table-1 for the period 1990-2013. Note these inputs are affecting both of the infants and their mothers and we expected that FPI, WS and GDP_{PPP} will have a negative impact on IMR while TFR will have a positive impact on IMR.

Table 1- Input and outputs data of the infant mortality rate model*.

Year	Fuzzy output	Crisp inputs			
(t)	$M_{<1}(t)$ ** $y^R = (\omega; v, \psi)_{LL}$	FPI	WS	TFR	GDP_{PPP}
1990	(42.3;1.6,12.1)	117.7	78.3	5.8	7442.8
1991	(41.6;1.6,11.9)	78.8	78.2	5.7	2691.9
1992	(40.8;1.6,11.7)	85.5	78.1	5.7	3548.7
1993	(40.2;1.6,11.5)	102.6	78.0	5.6	4595.8
1994	(39.5;1.5,11.3)	101.1	77.9	5.5	4729.1
1995	(38.9;1.5,11.2)	101.6	78.2	5.4	4781.2
1996	(38.3;1.5,11.0)	104.3	78.6	5.3	5239.7
1997	(37.6;1.5,10.8)	100.0	78.9	5.2	6261.9
1998	(37.0;1.4,10.5)	109.4	79.4	5.1	8274.3
1999	(36.4;1.4,10.4)	103.9	79.9	4.9	9581.2
2000	(35.8;1.4,10.3)	100.4	80.4	4.8	9647.8
2001	(35.2;1.4,10.1)	115.4	80.8	4.7	9810.5
2002	(34.6;1.3,9.9)	128.6	81.3	4.7	9019.1
2003	(34.1;1.3,9.8)	94.3	81.8	4.6	5989.1
2004	(33.5;1.3,9.6)	91.2	82.3	4.6	9237.9
2005	(33.0;1.3,9.5)	105.8	82.8	4.6	9697.9
2006	(32.4;1.2,9.3)	102.9	83.2	4.6	10733.4
2007	(31.8;1.2,9.1)	102.7	83.7	4.6	10893.2
2008	(31.2;1.2,8.9)	93.5	84.2	4.6	11715.8
2009	(30.6;1.2,8.8)	91.9	84.7	4.6	11875.0
2010	(30.0;1.2,8.6)	105.4	85.1	4.6	12417.7
2011	(29.3;1.1,8.4)	114.9	85.6	4.6	13203.0
2012	(28.7;1.1,8.2)	113.5	86.1	4.6	14813.5
2013	(27.9; 1.1,8.0)	126.0	86.5	4.6	15503.6

* The data sources of those shown in Table-1 are from the World Bank "http://data.worldbank.org/".
** Focusing our interest in symmetric QFN's the $M_{<1}(t)$ was fuzzified by researchers.

Often there is a problem to compare the effects of the inputs in which is the most influential for determining the output since the regression parameters depend on the units of measurement of the inputs as well as the dominance of their sizes. For this purpose, we need to get rid of units of measurements and the sizes of the inputs to make the comparison between them more meaningful. This can be done by standardizing the inputs in Table-1 using the equation $\left[(input_i - \overline{input})/\sigma_{input}\right]$, that mean we will measure the change of $M_{<1}(t)$ by the change of the standard deviation (σ_{input}) in each inputs. Comparison between the performances of the (12) and the method proposed by D'Urso [12] of the model (11) was made to know the difference when we use the position and entropy instead of membership function. Almost all the studies have been suggested so far using only triangular fuzzy numbers (i.e. R=1), but in this study we extended the AFLRM by using QFN with two orders (R=1,2) and three different sets of weights $\mathcal{W}_1 = (w_\xi = 1, w_s = 1, w_\mathcal{L} = w_\mathcal{R} = 1)$, $\mathcal{W}_2 = (w_\xi = 0.4, w_s = 0.1, w_\mathcal{L} = w_\mathcal{R} = 0.25)$ and

$\mathcal{W}_3 = (w_\xi = 0.6, w_s = 0.04, w_\mathcal{L} = w_\mathcal{R} = 0.18)$. The results have been tabulated below and the computing was all done using Matlab software.

Table 2- The estimates of parameters in models (11) and (12).

	Membership function M(1)	**R=1/Entropy (5) M(2)**	**R=2/Entropy (5) M(3)**	**R=1/Entropy (6) M(4)**	**R=2/Entropy (6) M(5)**
	$\Omega_1 = \begin{bmatrix} 33.6280 \\ -0.0708 \\ -2.7689 \\ -0.1162 \\ 1.3180 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$
	$\Omega_2 = \begin{bmatrix} 36.4303 \\ -0.0767 \\ -2.9996 \\ -0.1259 \\ 1.4278 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$
$\mathcal{W}_1$	$\lambda = \alpha = \begin{bmatrix} 1.1408e-4 \\ -0.2384 \\ 0.4969 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.1238e-4 \\ 0.1300 \\ 1.0788 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.4677e-4 \\ 0.0896 \\ 1.0924 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.1174e-4 \\ 0.1147 \\ 1.0661 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.4460e-4 \\ 0.0740 \\ 1.0741 \end{bmatrix}$
	$\Omega_1 = \begin{bmatrix} 33.6280 \\ -0.0708 \\ -2.7689 \\ -0.1162 \\ 1.3180 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$
	$\Omega_2 = \begin{bmatrix} 36.4303 \\ -0.0767 \\ -2.9996 \\ -0.1259 \\ 1.4278 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$
$\mathcal{W}_2$	$\lambda = \alpha = \begin{bmatrix} 2.2523e-4 \\ -0.3810 \\ 0.6286 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.7541e-4 \\ 0.1297 \\ 1.0830 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 2.5904e-4 \\ 0.0900 \\ 1.0878 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 1.7328e-4 \\ 0.1141 \\ 1.0726 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 2.5952e-4 \\ 0.0729 \\ 1.0878 \end{bmatrix}$
	$\Omega_1 = \begin{bmatrix} 33.6280 \\ -0.0708 \\ -2.7689 \\ -0.1162 \\ 1.3180 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$	$\theta_1 = \begin{bmatrix} 35.0291 \\ -0.0738 \\ -2.8843 \\ -0.1211 \\ 1.3729 \end{bmatrix}$
	$\Omega_2 = \begin{bmatrix} 36.4303 \\ -0.0767 \\ -2.9996 \\ -0.1259 \\ 1.4278 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$	$\theta_2 = \begin{bmatrix} 2.8023 \\ -0.0059 \\ -0.2307 \\ -0.0096 \\ 0.1098 \end{bmatrix}$
$\mathcal{W}_3$	$\lambda = \alpha = \begin{bmatrix} 6.9295e-4 \\ -0.7435 \\ 0.9632 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 2.5322e-4 \\ 0.1312 \\ 1.0642 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 5.5212e-4 \\ 0.0733 \\ 1.2962 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 2.5418e-4 \\ 0.1146 \\ 1.0667 \end{bmatrix}$	$\varphi = \tau = \begin{bmatrix} 5.0195e-4 \\ 0.0618 \\ 1.2266 \end{bmatrix}$

We can see from the estimated parameters in all models shown in Table-2 were achieved the expected relations between the inputs and the output, the estimated parameters of model (11) are shown in column M(1) where the fuzzification method was depending on membership function of the QFN while the others columns M(2), M(3), M(4) and M(5) are representing the estimated parameters of model (12) by the means of entropy function (3) and (4) with two different type of imprecision (R=1,2) respectively. The estimated parameters of Ω_1 and Ω_2 give an interval of the estimated parameters without fuzzification that interval represented by core but the estimated parameters of θ_1 represented the estimated parameters of data without fuzzification and θ_2 represented the length of uncertainty for each estimated parameter in θ_1. Notice the estimated parameters in λ (α) gave unusual relation between C_1 and the spread ψ because it seem little bit odd to have a negative relation between core and spread this was not an issue by the model we proposed, in particular in φ (τ), last but not least we can conclude that the weights were affecting the spreads and entropy only which is logically correct.

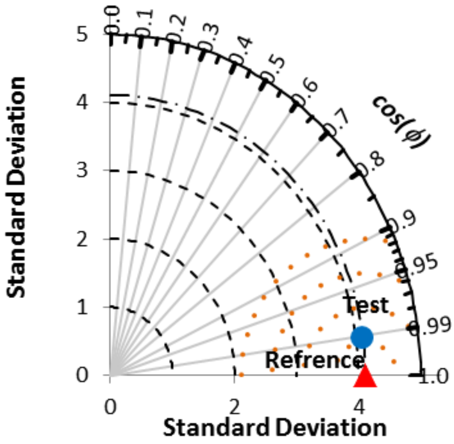
Moreover, as it turns out from the results reported in Table-2 we observe $M_{<1}(t)$ is strongly influenced by the access to improved water sources and the total fertility rate followed by the gross domestic product and least influenced by food production index.

Table 3- Evaluation Measures for Models (11) and (12)

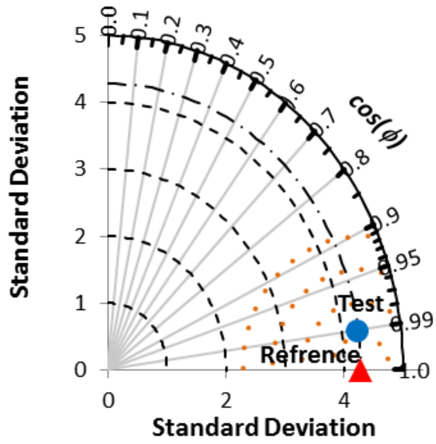
	Comparison Methods	Weights		
		$\mathcal{W}_1$	$\mathcal{W}_2$	$\mathcal{W}_3$
M(1)	$\mathcal{R}^2$ $(\mathcal{R}^2_{adjusted})$	0.9843 (0.9788)	0.9843 (0.9788)	0.9843 (0.9788)
	$\mathcal{AS}$	0.9563	0.9890	0.9874
	ALSD	1.0850	0.2035	0.2793
M(2)	$\mathcal{R}^2$ $(\mathcal{R}^2_{adjusted})$	0.9843 (0.9788)	0.9843 (0.9788)	0.9843 (0.9788)
	$\mathcal{AS}$	0.9776	0.9919	0.9888
	ALSD	0.6372	0.1880	0.2732
M(3)	$\mathcal{R}^2$ $(\mathcal{R}^2_{adjusted})$	0.9843 (0.9788)	0.9843 (0.9788)	0.9843 (0.9788)
	$\mathcal{AS}$	0.9773	0.9918	0.9887
	ALSD	0.6041	0.1847	0.2696
M(4)	$\mathcal{R}^2$ $(\mathcal{R}^2_{adjusted})$	0.9843 (0.9788)	0.9843 (0.9788)	0.9843 (0.9788)
	$\mathcal{AS}$	0.9779	0.9920	0.9888
	ALSD	0.6234	0.1866	0.2717
M(5)	$\mathcal{R}^2$ $(\mathcal{R}^2_{adjusted})$	0.9843 (0.9788)	0.9843 (0.9788)	0.9843 (0.9788)
	$\mathcal{AS}$	0.9776	0.9919	0.9888
	ALSD	0.5897	0.1833	0.2681

From Tabel 3- above we can see the using of the position and entropy of QFN did not change the statistical fit of models $\mathcal{R}^2(\mathcal{R}^2_{adjusted})$ at all, also the two different entropy function (3) and (4) almost gave an identical results in each type of imprecision (i.e. when R=1 or 2) but were superior the using of membership in model (11) as obvious from $\mathcal{AS}$ and ALSD measures.

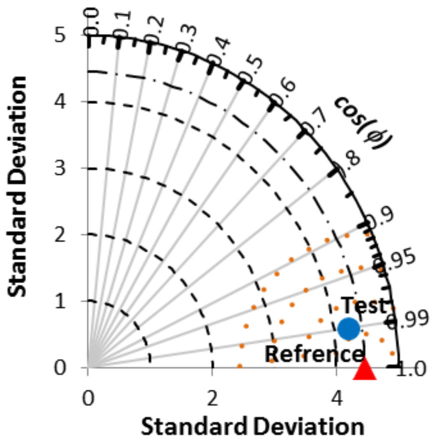
In order to visualize how well each estimated sub-models in models (11) and (14) fits the fuzzy observations we use Taylor diagram as showing in Figure-3 below the proposed model (14) was performs better by each sub-models except the position model and the C_1 model were have almost the same results (albeit it was slightly different).



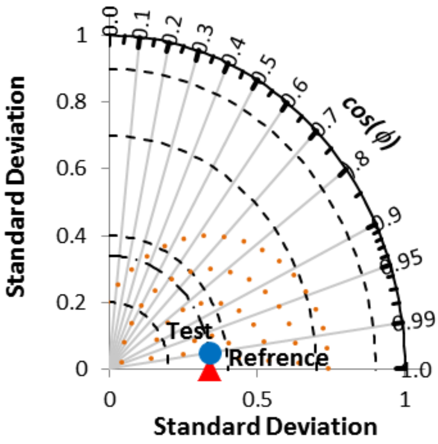
(a$_1$)



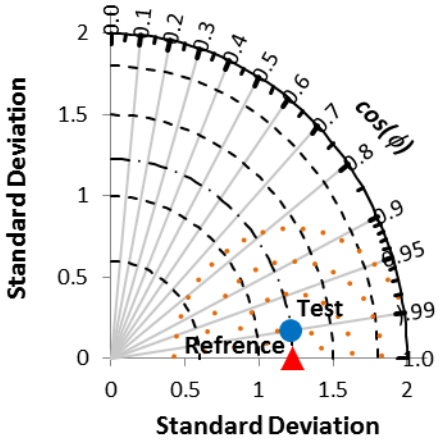
(a$_2$)



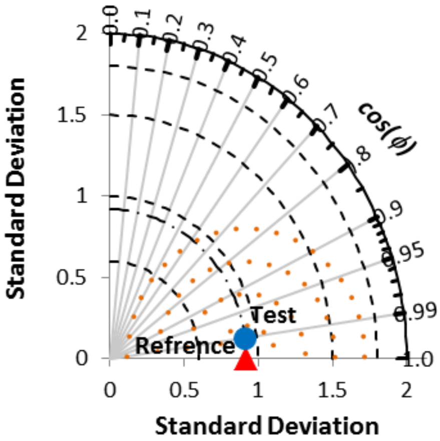
(b$_1$)



(b$_2$)



(c$_1$)



(c$_2$)

Figure 3- Taylor diagrams for sub-models in (11) represented by a$_1$,b$_1$ and c$_1$ (C_1, C_2 and ψ) and sub-models in (12) represented by a$_2$,b$_2$ and c$_2$ (ξ, s and e), red triangular represented the observation while the blue circle represented the estimated observation from the sub-models.

For completeness, we augment the estimation procedure by finding the estimated parameters of the implicit entropy model of the sub-models in (17), by applying the NNLS algorithm we obtained the following results as shown in Table-4 below:

Table 4- Estimated parameters of the implicit entropy model in (17).

M(2)	**M(3)**	**M(4)**	**M(5)**
$\beta_{\mathcal{L}} = \beta_{\mathcal{R}} = \begin{bmatrix} 6.7978 \\ 0 \\ 0 \\ 0 \\ 0.9030 \end{bmatrix}$	$\beta_{\mathcal{L}} = \beta_{\mathcal{R}} = \begin{bmatrix} 5.5624 \\ 0 \\ 0 \\ 0 \\ 0.7389 \end{bmatrix}$	$\beta_{\mathcal{L}} = \beta_{\mathcal{R}} = \begin{bmatrix} 6.2825 \\ 0 \\ 0 \\ 0 \\ 0.8346 \end{bmatrix}$	$\beta_{\mathcal{L}} = \beta_{\mathcal{R}} = \begin{bmatrix} 5.0260 \\ 0 \\ 0 \\ 0 \\ 0.6677 \end{bmatrix}$

Notice that in general the estimated intercept parameter shows a high degree of fuzziness (uncertainty) and the estimated parameter relative to fertility shows a low degree of fuzziness (uncertainty) whereas the estimated parameter relative to food production index, access to improved water sources and gross domestic product are null (there is no imprecision), maybe this procedure does not give a 100% certain estimates but exhibits the values of degree of fuzziness in model (12). However, the entropy functions M(5) had the least degree of fuzziness in estimate the parameters of entropy sub-model in model (17) comparing with the others.

5. Results Discussion

1. From Tables- 1 , 2 , 3 and Figure-3 we can conclude using the position with entropy function to describe fuzzy numbers is an efficient way and the model (12) was fitted pretty well whether using entropy (3) or (4) of QFN's and it superior the performance of model (11) with all sets of weights. However, The sets of arbitrary weights had affected and improved the performance of the estimation method especially the second one $\mathcal{W}_2$ which gave the highest $\mathcal{AS}$ with smallest ALSD.
2. Although the application in this paper was limited to some selective inputs which affect both of infant and their mothers, but one can draw to a close that to reduce infant mortality rate in Iraq we need to control women's fertility and increase the FPI, WS, GDP_{PPP} to each capita.

References

1. Wang, R. and Lu, J. **2009**. An enhanced fuzzy linear regression model with more flexible spreads, *Fuzzy Sets and Systems*, **160**: 2523 –2505.
2. D'Urso, P. and Gastaldi, T. **2000**. A least-squares approach to fuzzy linear regression analysis. *Computational Statistics & Data Analysis*, **34**: 427-440.
3. Coppi, R., D'Urso, P., Giordania, P. and Santorob, A. **2006**. Least squares estimation of a linear regression model with LR fuzzy response. *Computational Statistics & Data Analysis*, **51**: 286-267.
4. González, Rodrígueza G., Blancob, A., Colubib, A. and Asunción, Lubianob M. **2009**. Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets and Systems*, **160**: 357–370.
5. Dubois, D. and Prade, H. **1980**. *Fuzzy sets and systems: Theory and Applications*. Academic Press, New York.
6. Grzegorzewski P. and Mrówka E. **2005**. Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, **153**: 115–135.
7. Ferraro, M. B. **2008**. A linear regression model for LR fuzzy random variables: properties and inferential procedures. Department of Statistics, Probability and Applied statistics, Faculty of Statistics, Sapienza University of Rome. Piazzale Aldo Mor, Rome, Italy.
8. Kao G. and Lin P-H. **2005**. Entropy for fuzzy regression analysis. *International Journal of ystems Science*, **36**(15): 869-876 .
9. De, Luca A. and Termini, S. **1972**. A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, **20**(4): 301-312.
10. Wang, W-G. and Chiu, C-H. **1999**. The entropy change in extension principle. *Fuzzy Sets and System*, **103**: 153-162.
11. D'Urso, P. and Giordani, P. **2006**. A weighted fuzzy c-means clustering model for fuzzy data. *Computational Statistics & Data Analysis*, **50**:1496–1523.
12. D'Urso, P. **2003**. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. *Computational Statistics & Data Analysis*, **42**: 47 – 72.

13. Chen, D. and Plemmons , R. **2007**. Nonnegativity constraints in numerical analysis, in: Bultheel A. and Cools R. (Editors), Conference Proceedings of the Symposium of the Birth of Numerical Analysis. Leuven, Belgium, 29-30 October.

14. Taylor K. **2001**. Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research*, **106**(D7): 7183-7192.

15. Zakir, M. and Wunnava, P. **1999**. Factors affecting infant mortality rates: evidence from cross–sectional data. *Applied Economics Letters*, **6**: 271-273.

16. Szymański, A. and Rossa, A. **2014**. Fuzzy Mortality Model Based on Banach Algebra. *International Journal of Intelligent technologies and applied statistics*, **7**(3): 241-265.