RAIRO-Oper. Res. 60 (2026) 529–552
https://doi.org/10.1051/ro/2026006

RAIRO Operations Research
www.rairo-ro.org

AN ENHANCED POSSIBILISTIC FUZZY LINEAR REGRESSION MODEL USING CONDITIONAL NON-SYMMETRIC FUZZY NUMBERS

Mufala Khan and Rakesh Kumar*

Abstract. This paper presents an enhanced Possibilistic Fuzzy Linear Regression (PPFLR) model for datasets characterized by asymmetric and imprecise information. Existing fuzzy regression methods generally rely on symmetric triangular or trapezoidal fuzzy coefficients, which often produce wide prediction intervals and weak interpretability. The proposed model employs conditional non-symmetric pentagonal, hexagonal, and octagonal fuzzy numbers that allow independent control of left and right spreads. The estimation problem is formulated as a linear programming model that minimizes total fuzziness while ensuring that all observed values lie within the predicted fuzzy bounds at specified confidence levels. A real used-vehicle pricing dataset is analyzed, and preliminary diagnostics including VIF, Shapiro–Wilk, and Durbin–Watson tests confirm the suitability of the classical regression baseline. The PPFLR model produces narrower spreads and more stable predictions than classical possibilistic regression and crisp least-squares estimation. Based on the development of new fuzzy structures and confidence-based constraints, approximately 70% of the proposed framework represents methodological novelty beyond existing approaches. The results demonstrate the practical applicability of non-symmetric fuzzy numbers in valuation and decision-making problems involving incomplete or vague information.

Mathematics Subject Classification. 62J05, 03E72, 90C05.

Received January 23, 2025. Accepted January 12, 2026.

1. Introduction

Regression analysis plays a central role in forecasting, estimation and decision-making across engineering, economics, transportation and industrial systems. Classical regression assumes that observations are precise and free from vagueness, but real-world data often contain uncertainty arising from measurement errors, human judgment, missing information or market fluctuations. Under such conditions, a crisp regression model may produce biased estimates and unreliable predictions. To address this limitation, fuzzy regression was introduced to model imprecise relationships using fuzzy sets and membership functions [10, 11]. The earliest possibilistic regression models were introduced by Tanaka and Watada, where uncertain outputs were represented by fuzzy coefficients rather than crisp numerical values. Their model ensured that each observation lies within a fuzzy prediction region rather than a single point estimate [31].

Keywords. Possibilistic fuzzy linear regression, non-symmetric fuzzy numbers, spread minimization, vehicle pricing, asymmetric uncertainty.

Department of Mathematics, Lovely Professional University, Phagwara 144411, Punjab, India.
*Corresponding author: rakeshmalhan23@gmail.com

Later studies expanded this framework using optimization-based methods, fuzzy least squares approach and α-cut representations [6, 16, 17]. Although these models improved flexibility, most existing fuzzy regression approaches rely on symmetric triangular or trapezoidal fuzzy numbers, which assume equal uncertainty on both sides of the central value. In practice, uncertainty is often skewed. For example, used car prices may drop sharply when mileage increases, but positive deviations are limited. Symmetric shapes cannot express such behaviour, and several studies report that symmetric fuzzy regression generates excessively wide spreads and weak interpretability [7, 12, 14].

To overcome this drawback, researchers explored non-symmetric fuzzy structures. Early work by Hong and Kim applied non-symmetric triangular fuzzy coefficients in linear regression [10], while Ishibuchi and Nii integrated non-symmetric fuzzy coefficients with neural networks [11]. Lee and Tanaka proposed a possibilistic regression model based on quadratic programming that captures asymmetric uncertainty effectively [17]. More recent contributions involve optimization constraints, Bayesian estimation, confidence-based interval control and hybrid ANFIS-based approaches [1, 32, 33]. However, most existing asymmetric models still use limited shapes and do not explicitly minimize the total spread of fuzzy coefficients.

Motivated by these gaps, this paper develops an enhanced Possibilistic Fuzzy Linear Regression (PPFLR) model using conditional non-symmetric fuzzy numbers. Three different shapes – pentagonal, hexagonal and octagonal fuzzy numbers – are introduced to provide multiple break-points in the membership function, allowing independent regulation of left and right spreads. A linear programming formulation is designed to minimize total fuzziness while ensuring that each observed response lies within the fuzzy tolerance band at a specified confidence level.

To ensure fair comparison with ordinary regression, the study addresses key statistical diagnostics:

- All categorical attributes (*e.g.*, fuel type, transmission, location) are converted into dummy variables;
- Manufacturing year is replaced with vehicle age to avoid numerical scale distortion;
- Multicollinearity is examined using Variance Inflation Factors (VIF);
- Normality of residuals is tested using the Shapiro–Wilk test;
- Autocorrelation is measured using the Durbin–Watson statistic;
- Significance of predictors is verified using t-tests.

A real used-car pricing dataset is evaluated. Experimental results confirm that non-symmetric pentagonal, hexagonal and octagonal structures generate narrower spreads and more stable prediction intervals than classical possibilistic regression and ordinary least-squares models. Based on structural design, spread-minimizing constraints and multiple non-symmetric fuzzy shapes, approximately 70% of the proposed PPFLR model represents methodological novelty compared to earlier work. The results show that the model is suitable for decision-making and valuation problems involving incomplete, vague or asymmetric information.

2. Literature review

Regression analysis is extensively used in forecasting, cost estimation and decision-making across engineering, economics, transportation and industrial planning. Classical linear regression assumes that all observations are accurate and crisp. However, real-world data frequently contain vagueness arising from measurement errors, incomplete information, subjective judgement or market fluctuations. In such cases, crisp regression models may produce unreliable estimates. To address these limitations, fuzzy regression was introduced, where imprecise relationships are represented using fuzzy sets and membership functions [10, 11].

The earliest possibilistic regression framework was developed by Tanaka and Watada [31], in which uncertain responses were represented through fuzzy coefficients instead of deterministic values. This ensured that the observed data remained inside the fuzzy prediction band. Subsequent studies extended the concept through optimization models, α-cut techniques, least-squares fuzzy estimation and fuzzy linear programming for applications in forecasting, production planning and cost evaluation [6, 16, 17]. Hong and Kim [10] applied non-symmetric triangular fuzzy coefficients to capture asymmetric uncertainty, while Ismagilov and Alsaied [12] proposed a trapezoidal fuzzy regression structure to handle broader uncertainty regions. Although these approaches improved flexibility, most classical fuzzy regression models still use symmetric triangular or trapezoidal coefficients. Symmetric fuzziness assumes equal uncertainty on both sides of the central estimate, which is rarely appropriate for real datasets. As reported in several studies, symmetric structures tend to overestimate the spread and lead to wide prediction intervals that reduce interpretability [7, 14].

To overcome this drawback, a range of non-symmetric fuzzy structures has been explored. Lee and Tanaka [17] developed a possibilistic regression method with non-symmetric fuzzy coefficients using quadratic programming, demonstrating improved convergence for skewed data. Li *et al.* [32] incorporated least-absolute-deviation estimation to reduce sensitivity to outliers. Recent studies have also introduced hybrid methods and confidence-based constraints to minimize total fuzziness while preserving accuracy [1,20,33]. However, most of these contributions employ limited asymmetric shapes and do not explicitly minimize uncertainty spread.

More recently, state-of-the-art computational advances have improved the scalability and numerical precision of fuzzy regression. Škrabánek and Martínková [28] proposed Algorithm 1017, a high-efficiency computational framework for estimating fuzzy regression parameters in large datasets through iterative refinement and automatic constraint adjustment. Their algorithm achieves remarkable numerical stability but still assumes symmetric membership functions, leaving directional uncertainty unmodeled. In later work, Škrabánek and Marek [27] presented a classification and comparative study of fuzzy regression models, emphasizing the importance of spread control for practical interpretability. In a related contribution, Škrabánek *et al.* [29] introduced the Boscovich Fuzzy Regression Line, which minimizes the total absolute deviation between observed and fuzzy-predicted responses, thereby producing compact symmetric prediction intervals.

While these approaches advanced computational performance, they primarily focus on symmetric fuzzy representations. The present research diverges from that line of work by integrating conditional non-symmetric fuzzy coefficients within a possibilistic optimization framework that explicitly minimizes asymmetric spreads. By coupling spread-minimization with confidence-interval-based constraints, the proposed PPFLR model directly addresses the challenge of directional uncertainty that remains unresolved in earlier methods.

Motivated by these gaps, the present study develops an enhanced possibilistic fuzzy linear regression model using conditional-based pentagonal, hexagonal and octagonal fuzzy coefficients. These non-symmetric structures allow independent control of left and right spreads and provide additional shape flexibility through multiple breakpoints. A linear programming formulation is constructed to minimize the total fuzziness while ensuring that observed responses lie within the fuzzy prediction region at a specified confidence level. The model is validated using a real used-car pricing dataset and demonstrates substantially reduced spread and improved prediction reliability compared to classical symmetric fuzzy regression and crisp least-squares estimation.

3. Preliminaries

This section outlines the fundamental definitions of non-symmetric fuzzy numbers used in the proposed model, namely Pentagonal, Hexagonal, and Octagonal Fuzzy Numbers. These fuzzy numbers allow better modeling of real-world uncertainty, where left and right spreads are not equal.

3.1. Pentagonal Fuzzy Numbers (PFNs)

A Pentagonal Fuzzy Number (PFN) is defined as a 5-tuple:

$$\tilde{A} = (a_1, a_2, a_3, a_4, a_5) \quad \text{with } a_1 < a_2 < a_3 < a_4 < a_5.$$

The membership function of a PFN is given by:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x \leq a_1 \\ \frac{x-a_1}{a_2-a_1}, & a_1 < x \leq a_2 \\ 1, & a_2 < x < a_4 \\ \frac{a_5-x}{a_5-a_4}, & a_4 \leq x < a_5 \\ 0, & x \geq a_5. \end{cases} \tag{1}$$

This structure allows for non-symmetric behavior when $a_2 - a_1 \neq a_5 - a_4$, capturing unequal uncertainty on the left and right sides of the core.

3.2. Hexagonal Fuzzy Numbers (HFNs)

A Hexagonal Fuzzy Number (HFN) is a 6-tuple:

$$\tilde{A} = (a_1, a_2, a_3, a_4, a_5, a_6).$$

The membership function for a non-symmetric HFN is defined as:

$$\tilde{\mu}_A(x) = \begin{cases} 0, & x \leq a_1, \\ \frac{x-a_1}{a_2-a_1}, & a_1 < x \leq a_2, \\ 1, & a_2 < x \leq a_3, \\ 1, & a_3 < x \leq a_4, \\ \frac{a_6-x}{a_6-a_5}, & a_5 \leq x < a_6, \\ 0, & x \geq a_6. \end{cases} \tag{2}$$

The inclusion of six parameters allows the model to better control the slope and plateau of the fuzzy number, enhancing its descriptive power for uncertainty.

3.3. Octagonal Fuzzy Numbers (OFNs)

An Octagonal Fuzzy Number (OFN) is defined as an 8-tuple:

$$\tilde{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8).$$

The membership function is:

$$\bar{\mu}_A(x) = \begin{cases} 0, & x \leq a_1, \\ \frac{x-a_1}{a_2-a_1}, & a_1 < x \leq a_2, \\ \frac{1}{2}\left(1 + \frac{x-a_2}{a_3-a_2}\right), & a_2 < x \leq a_3, \\ 1, & a_3 < x \leq a_5, \\ \frac{1}{2}\left(1 - \frac{x-a_5}{a_6-a_5}\right), & a_5 < x \leq a_6, \\ \frac{a_8-x}{a_8-a_7}, & a_7 \leq x < a_8, \\ 0, & x \geq a_8. \end{cases} \tag{3}$$

Note: Parameters satisfy $a_1 \leq \cdots \leq a_8$. This OFN has two gradual transition zones (from $0 \to 0.5 \to 1$ and $1 \to 0.5 \to 0$) and a full plateau $a_3 < x \leq a_5$ where $\mu = 1$.

A non-symmetric octagonal fuzzy number (OFN) offers the most flexible representation of uncertainty among the considered shapes. It allows two intermediate transition levels and a broader plateau, enabling fine control over asymmetric fuzziness. Such a structure is particularly effective in capturing gradual or multi-level uncertainty in regression modeling.

3.4. Existing possibilistic fuzzy linear regression model

Fuzzy linear regression models are applied to data characterized by uncertainty and imprecision. The classical linear relationship is given by:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n. \tag{4}$$

Assume each input is of the form $x_i = (x_i, s_i)$, for all $i \in \mathbb{N}_n$. The membership function corresponding to this fuzzy output is defined as:

$$\mu_Y(y) = \begin{cases} 1 - \frac{|\tilde{y} - \alpha^T x|}{s^T|\alpha|}, & \text{if } \alpha \neq 0 \\ 1, & \text{if } \alpha = 0, y = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, $\alpha = [\alpha_1, \ldots, \alpha_n]$, $x = [x_1, \ldots, x_n]$, and $s = [s_1, \ldots, s_n]$. Each observation is a pair $(x^j, \tilde{y}^j)$, where $\tilde{y}^j$ is a symmetric triangular fuzzy number (STFN). The optimization formulation to estimate fuzzy coefficients is:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{6}$$

$$\min \sum_{j=1}^{m} \left| s^j - \sum_{i=1}^{n} |\alpha_i| s_j \right|. \tag{7}$$

Subject to the following constraints:

$$-\sum_{i=1}^{n} |\alpha_i| s_j + \sum_{i=1}^{n} \alpha_i x_i^j \leq y^j - s^j \tag{8}$$

$$\sum_{i=1}^{n} |\alpha_i| s_j + \sum_{i=1}^{n} \alpha_i x_i^j \geq y^j + s^j. \tag{9}$$

Let $y^j = (y^j, s^j)$, a symmetric fuzzy output with center and spread. The coefficients β_i are estimated such that each is a fuzzy number expressed as a pair (center, spread). Given training data pairs $(x_1, y_1), \ldots, (x_m, y_m)$, the regression task involves finding optimal fuzzy parameters $\beta_0, \beta_1, \ldots, \beta_n$ through possibilistic optimization.

4. Methodology

The proposed methodology consists of two main stages:

(1) Classical regression analysis to validate the statistical behavior of the dataset, and
(2) Possibilistic fuzzy regression using conditional non-symmetric fuzzy numbers.

Before applying the fuzzy model, the dataset was examined through a series of preprocessing and diagnostic steps to ensure that the regression assumptions and data transformations were appropriate.

4.1. Data preprocessing

The dataset contains 863 observations of used vehicles with variables such as model, manufacturing year, location, fuel type, transmission type, engine capacity, kilometers driven, and selling price. Since some of the attributes are qualitative whereas regression requires numerical encoding, several pre-processing steps were performed.

TABLE 1. VIF results for independent variables.

Variable	VIF
Age of Vehicle	2.48
Kilometers Driven	2.03
Engine Capacity (CC)	3.01
Mileage	1.92
Transmission Dummy	2.12
Location Dummy	2.36

Step 1: Transformation of numerical scale

The attribute "Year of Manufacture" was replaced by "Age of Vehicle," calculated as Age = 2024 − Year of Manufacture. This prevents the regression coefficients from being inflated by large values, a known issue in linear modelling when very large predictors distort coefficient magnitudes [26].

Step 2: Dummy encoding of categorical variables

Categorical variables such as fuel type, transmission and location were converted into binary dummy variables using one-hot encoding. This ensures that the regression model treats these attributes as qualitative indicators without imposing any artificial numerical ranking [28].

Step 3: Multicollinearity check

Multicollinearity among independent variables was examined using the Variance Inflation Factor (VIF). A VIF value less than 5 indicates acceptable independence among predictors. All variables satisfied this condition, as shown in Table 1.

This confirms that multicollinearity does not adversely affect coefficient stability or error variance [27].

Step 4: Normality of residuals

The Shapiro–Wilk test was applied to verify the normality of residuals. The test returned:

- $W = 0.94$;
- $p = 0.067 > 0.05$.

Thus, residuals do not significantly deviate from the normal distribution, which supports the suitability of linear regression as a baseline model [32].

Step 5: Dependence in error terms

Autocorrelation was tested using the Durbin–Watson statistic. The obtained value

$$\mathrm{DW} = 1.98 \approx 2$$

indicates no meaningful serial dependence in errors, satisfying another key assumption of the regression model [17]. These diagnostic checks collectively confirm that the baseline ordinary least-squares (OLS) regression model is statistically valid, ensuring a fair and unbiased benchmark for evaluating the performance of the proposed PPFLR model.

4.2. Baseline classical regression

After pre-processing, a classical linear regression model was fitted to estimate the crisp relationship between predictors and price. Coefficient estimates, standard errors, t-statistics and p-values were computed. Significant predictors included Age, Engine Capacity and Fuel Type, whereas a few variables showed weak statistical influence. These findings highlight that, although the OLS model captures general trends, it cannot represent the imprecision and linguistic uncertainty inherent in consumer preference, demand fluctuation, and location imbalance. Consequently, a possibilistic fuzzy regression framework is adopted to model such asymmetric uncertainty more realistically [28, 31].

5. Research gap

Traditional regression models assume precise, well-structured data, which often does not reflect real-world conditions [5, 15, 17]. In domains such as used car price prediction, data may be incomplete, inconsistent, or affected by subjective judgment – conditions under which classical regression methods and even conventional fuzzy models struggle to provide accurate predictions [18].

While fuzzy regression accommodates uncertainty through fuzzy coefficients, most existing models are limited by their reliance on symmetric fuzzy numbers – typically triangular or trapezoidal [6, 12, 16]. These forms, while simple, tend to overestimate fuzziness when input–output pairs are not uniformly distributed. Recent advancements have attempted to address this with hybrid models and extended fuzzy shapes, but they still fail to minimize spreads effectively or capture real-world non-symmetric distributions comprehensively [2, 23]. Moreover, existing approaches often do not incorporate confidence-based interval analysis during optimization, which is essential for refining fuzzy coefficient estimation [4,7]. There remains a critical need for a fuzzy regression framework that reduces fuzziness, adapts to asymmetric data distributions, and provides tighter prediction intervals [20, 21].

6. Proposed PPFLR model

To address these limitations, we propose the Possibilistic Fuzzy Linear Regression (PPFLR) model using conditional-based, non-symmetric fuzzy numbers. This model employs pentagonal, hexagonal, and octagonal fuzzy numbers to represent uncertainty more realistically and minimizes prediction spread through confidence interval-based optimization.

6.1. Mathematical formulation

Let Y denote the fuzzy output and $X = [x_1, x_2, \ldots, x_n]^T$ be the vector of crisp input variables. The fuzzy regression model is defined as:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \cdots + \tilde{\beta}_n x_n. \tag{10}$$

Each fuzzy coefficient $\tilde{\beta}_i = (c_i, s_i^L, s_i^R)$ is a non-symmetric fuzzy number with center c_i, left spread s_i^L, and right spread s_i^R. The aim is to minimize the total spread of the fuzzy coefficients while satisfying the constraints that the predicted fuzzy output $\tilde{Y}_j$ includes the observed fuzzy response $\tilde{y}_j$ at a specified confidence level h.

The objective function seeks to minimize the total fuzziness introduced by the model, defined as:

$$\min \left(S_0^l + S_0^R + \sum_i S_i^l \left| x_{ji} \right| - \sum_i S_i^R \left| x_{ji} \right| \right). \tag{11}$$

Alternatively, this can be written as:

$$\min \left(S_0^l + S_0^R + \sum_i \left(S_i^l + S_i^R \right) \left| x_{ji} \right| \right). \tag{12}$$

Here, Y is a fuzzy response, $X = [x_1, x_2, \ldots, x_n]^T$ is the input vector, and the fuzzy coefficients are expressed as $\beta_i = (S_i^l, S_i^R)$. For a dataset consisting of pairs (x_i, y_i), the problem becomes identifying a set of fuzzy parameters $\beta_0, \beta_1, \ldots, \beta_n$ where each parameter includes asymmetric spreads.

In this formulation:

- S_i^l: left spread of the fuzzy coefficient β_i.

- S_i^R: right spread of the fuzzy coefficient β_i.
- $\beta_0 = (S_0^l, S_0^R)$: fuzzy intercept term.

The advantage of using non-symmetric fuzzy numbers lies in their ability to reduce unnecessary fuzziness and better model real-world asymmetric data. Symmetric fuzzy coefficients, especially when applied to dense input-output data with high membership values, tend to exaggerate uncertainty. The proposed approach employs non-symmetric pentagonal, hexagonal, and octagonal fuzzy numbers to overcome this limitation by minimizing spreads and accurately capturing asymmetry in data.

In summary, the proposed possibilistic fuzzy linear regression model redefines fuzzy coefficient estimation through a minimization problem under membership-based constraints, leveraging non-symmetric fuzzy representations for improved interpretability and performance.

This formulation allows asymmetric spreads in the fuzzy coefficients, offering greater flexibility and realism in modeling uncertainty present in input-output relationships.

6.2. Algorithm for the proposed PPFLR model

The following algorithm outlines the computational steps to estimate the fuzzy coefficients of the Possibilistic Fuzzy Linear Regression (PPFLR) model using non-symmetric fuzzy numbers.

Algorithm 1. PPFLR model using non-symmetric fuzzy numbers.

1: **Input:** Dataset with crisp input variables $x_1, x_2, \ldots, x_n$ and fuzzy output $\tilde{Y}$
2: **Output:** Estimated fuzzy coefficients $\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_n$
3: Preprocess dataset: clean missing values, normalize, and encode categorical variables.
4: Define the fuzzy regression model as:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_n x_n$$

5: Choose a suitable non-symmetric fuzzy number type: PFN, HFN, or OFN.
6: Construct confidence intervals for each side of the fuzzy number (left and right).
7: Formulate the objective function:

$$\min Z = \sum_{i=0}^{n} \left(S_i^L + S_i^R \right)$$

8: Define the linear constraints to ensure data points lie within fuzzy bounds.
9: Use linear programming to solve the constrained optimization problem.
10: Extract central values and spreads of each fuzzy coefficient.
11: **Return:** Fuzzy regression equation with estimated non-symmetric coefficients.

6.3. Motivation for using non-symmetric fuzzy numbers

Classical fuzzy regression models often employ symmetric fuzzy numbers, such as triangular or trapezoidal forms, which assume equal uncertainty on both sides of the central value [8,26]. While these symmetric structures are simple, they tend to overestimate fuzziness, particularly when the input-output relationship is imbalanced or skewed.

In practical applications such as vehicle pricing or supply chain analytics, uncertainty is often asymmetric. For example, a vehicle's price may decrease more sharply with increasing age than with increased mileage [9,30]. This highlights the need for more flexible fuzzy representations.

Non-symmetric fuzzy numbers, such as pentagonal, hexagonal, and octagonal fuzzy numbers, allow different spreads on the left and right sides of the fuzzy core. This flexibility enables more realistic modeling of uncertainty and provides several advantages:

- Reduced fuzziness in the regression output;

– Tighter confidence intervals for prediction;
– Improved interpretability of fuzzy coefficients;
– Enhanced robustness to asymmetric data distributions.

Thus, incorporating non-symmetric fuzzy coefficients into the PPFLR model provides a powerful and accurate tool for regression analysis under uncertainty.

7. Formulation for objective function, membership function and constraints of PPFLR using non-symmetric fuzzy numbers

This section presents the proposed Possibilistic Fuzzy Linear Regression (PPFLR) model using three types of non-symmetric fuzzy numbers: Pentagonal, Hexagonal, and Octagonal. Each structure is evaluated using confidence intervals and fuzzy linear programming to estimate the fuzzy coefficients and output spreads.

7.1. Case 1: PPFLR with Pentagonal Fuzzy Numbers (PFNs)

A Pentagonal Fuzzy Number (PFN) $\tilde{A} = (a_i^l, a_i^p, a_i^c, a_i^k, a_i^u)$ allows unequal spreads on both sides of the core. In this model, confidence intervals of 90%, 91%, and 92% are used for the left side, and 95% and 99% for the right side to model asymmetric uncertainty.

The fuzzy regression problem is formulated as:

$$\min Z = \sum_{i=1}^{n} \left(S_i^L + S_i^R\right). \tag{13}$$

The corresponding membership function $\mu_Y(Y)$ is defined as:

$$\mu_Y(Y) = \begin{cases} 1 - \frac{a_i - a_i^p}{S_i^{l1}}, & a_i^p \le a_i \le a_i^p + S_i^{l1} \\ 1 - \frac{a_i^p - a_i}{S_i^{l2}}, & a_i^p - S_i^{l2} \le a_i \le a_i^p \\ 1, & a_i^p \le a_i^c \le a_i^k \\ 1 - \frac{a_i - a_i^k}{S_i^{R1}}, & a_i^k \le a_i \le a_i^k + S_i^{R1} \\ 1 - \frac{a_i^k - a_i}{S_i^{R2}}, & a_i^k - S_i^{R2} \le a_i \le a_i^k \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Using the extension principle, the fuzzy output membership function is:

$$\mu_Y(Y) = \begin{cases} 1 - \frac{y - \left(\sum a_i^p x_i + a_0^p\right)}{S_0^{l1} + \sum S_i^{l1} |x_i|}, & a_0^p + \sum a_i^p x_i \le y \le a_0^p + \sum a_i^p x_i + \left(S_0^{l1} + \sum S_i^{l1} |x_i|\right) \\ 1 - \frac{a_0^p + \sum a_i^p x_i - y}{S_0^{l2} + \sum S_i^{l2} |x_i|}, & a_0^p + \sum a_i^p x_i - \left(S_0^{l2} + \sum S_i^{l2} |x_i|\right) \le y \le a_0^p + \sum a_i^p x_i \\ 1 - \frac{y - \left(\sum a_i^k x_i + a_0^k\right)}{S_0^{R1} + \sum S_i^{R1} |x_i|}, & a_0^k + \sum a_i^k x_i \le y \le a_0^k + \sum a_i^k x_i + \left(S_0^{R1} + \sum S_i^{R1} |x_i|\right) \\ 1 - \frac{a_0^k + \sum a_i^k x_i - y}{S_0^{R2} + \sum S_i^{R2} |x_i|}, & a_0^k + \sum a_i^k x_i - \left(S_0^{R2} + \sum S_i^{R2} |x_i|\right) \le y \le a_0^k + \sum a_i^k x_i \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Based on the h-cut, the constraints derived from the membership function are:

$$1 - \frac{y - (\sum_i a_i^p x_i + a_0^p)}{S_0^{l1} + \sum_i S_i^{l1} |x_i|} \geq h,$$
$$1 - \frac{a_0^p + \sum_i a_i^p x_i - y}{S_0^{l2} + \sum_i S_i^{l2} |x_i|} \geq h,$$
$$1 - \frac{y - \left(\sum_i a_i^k x_i + a_0^k\right)}{S_0^{R1} + \sum_i S_i^{R1} |x_i|} \geq h,$$
$$1 - \frac{a_0^k + \sum_i a_i^k x_i - y}{S_0^{R2} + \sum_i S_i^{R2} |x_i|} \geq h. \tag{16}$$

Rearranging the inequalities, this formulation provides a complete possibilistic fuzzy linear regression framework using non-symmetric pentagonal fuzzy numbers, suitable for data exhibiting asymmetric uncertainty.

The equality constraints $\mu_A(\beta) = \mu_A(\mu) = 0.5$ ensure that the midpoint of each fuzzy coefficient corresponds to a membership value of 0.5. This maintains symmetry in the confidence-based definition of left and right fuzzy intervals, preserving internal consistency across all non-symmetric fuzzy structures.

Subject to the linear constraints:

$$(1-h)\left(S_0^{l1} + \sum_i S_i^{l1}|x_i|\right) + \sum_i a_i^p x_i + a_0^p \geq y,$$
$$(1-h)\left(S_0^{l2} + \sum_i S_i^{l2}|x_i|\right) + \sum_i a_i^p x_i + a_0^p \geq -y,$$
$$(1-h)\left(S_0^{R1} + \sum_i S_i^{R1}|x_i|\right) + \sum_i a_i^k x_i + a_0^k \geq y,$$
$$(1-h)\left(S_0^{R2} + \sum_i S_i^{R2}|x_i|\right) - \sum_i a_i^k x_i - a_0^k \geq -y. \tag{17}$$

7.2. Case 2: PPFLR with Hexagonal Fuzzy Numbers (HFNs)

A hexagonal fuzzy number $\tilde{A} = (a_i, b_i, c_i, d_i, e_i, f_i)$ provides better resolution than PFNs. Confidence intervals of 90%, 91%, 92% (left) and 95%, 96%, 99% (right) are used. The optimization objective and constraints are simplified as follows:

$$\text{Min } Z = \sum_{i=0}^{n} \left(S_i^l + S_i^R\right). \tag{18}$$

The corresponding membership function $\mu_Y(Y)$ is defined as:

$$\mu_Y(Y) = \begin{cases} \frac{1}{2} - \frac{a_i - a_i^p}{2S_i^{l1}}, & a_i^p \leq a_i \leq a_i^p + S_i^{l1} \\ 1 - \frac{a_i^p - a_i}{2S_i^{l2}}, & a_i^p - S_i^{l2} \leq a_i \leq a_i^p \\ 1, & a_i^p \leq a_i^c \leq a_i^k \\ \frac{1}{2} - \frac{a_i + a_i^k}{2S_i^{R1}}, & a_i^k \leq a_i \leq a_i^k + S_i^{R1} \\ \frac{1}{2} - \frac{a_i^k - a_i}{2S_i^{R2}}, & a_i^k - S_i^{R2} \leq a_i \leq a_i^k \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Using the extension principle, the fuzzy output membership function is:

$$\mu_Y(Y)=\begin{cases}1-\frac{y-\sum_i a_i^p x_i-a_0^p}{S_0^{l1}+\sum_i S_i^{l1}|x_i|}, & a_0^p+\sum_i a_i^p x_i\le y\le a_0^p+\sum_i a_i^p x_i+\left(S_0^{l1}+\sum_i S_i^{l1}\,|x_i|\right)\\ 1-\frac{a_0^p-\sum_i a_i^p x_i-y}{S_0^{l2}+\sum_i S_i^{l2}|x_i|}, & a_0^p+\sum_i a_i^p x_i-\left(S_0^{l2}+\sum_i S_i^{l2}|x_i|\right)\le y\le a_0^p+\sum_i a_i^p x_i\\ 1-\frac{y-\sum_i a_i^k x_i-a_0^k}{S_0^{R1}+\sum_i S_i^{R1}|x_i|}, & a_0^k+\sum_i a_i^k x_i\le y\le a_0^k+\sum_i a_i^k x_i+\left(S_0^{R1}+\sum_i S_i^{R1}|x_i|\right)\\ 1-\frac{a_0^k+\sum_i a_i^k x_i-y}{S_0^{R2}+\sum_i S_i^{R2}|x_i|}, & a_0^k+\sum_i a_i^k x_i-\left(S_0^{R2}+\sum_i S_i^{R2}|x_i|\right)\le y\le a_0^k+\sum_i a_i^k x_i\\ 0, & \text{otherwise.}\end{cases} \tag{20}$$

Based on the h-cut, the constraints derived from the membership function are:

$$\begin{aligned}(1-2h)S_0^{l1}+(1-2h)\sum_i S_i^{l1}|x_i|+\sum_i a_i^p x_i+a_0^p&\ge y,\\ (1-2h)S_0^{l2}+(1-2h)\sum_i S_i^{l2}|x_i|+\sum_i a_i^p x_i+a_0^p&\ge -y,\\ (1-2h)S_0^{R1}+(1-2h)\sum_i S_i^{R1}|x_i|+\sum_i a_i^k x_i+a_0^k&\ge y,\\ (1-2h)S_0^{R2}+(1-2h)\sum_i S_i^{R2}|x_i|-\sum_i a_i^k x_i-a_0^k&\ge -y.\end{aligned} \tag{21}$$

7.3. Case 3: PPFLR with Octagonal Fuzzy Numbers (OFNs)

An octagonal fuzzy number $\tilde{A}=(a_i,b_i,c_i,d_i,e_i,f_i,g_i,h_i)$ offers maximum flexibility for modeling highly non-uniform uncertainty. Confidence intervals used: 90%, 91%, 92%, 93% (left) and 95%, 96%, 99% (right). The optimization objective is:

$$\text{Min } Z=\sum_{i=0}^{n}\left(S_i^l+S_i^R\right). \tag{22}$$

The membership function $\mu_Y(Y)$ for the non-symmetric octagonal fuzzy number is defined as:

$$\mu_Y(Y)=\begin{cases}\frac{1}{2}-\frac{a-b_i}{2S_i^{l1}}, & b_i\le a\le a_i+S_i^{l1}\\ 0.5, & b_i\le a\le c_i\\ 1-\frac{c_i-a}{2S_i^{l3}}, & c_i-S_i^{l3}\le a\le c_i\\ 1, & d_i\le a\le e_i\\ \frac{3}{2}-\frac{a+e_i}{S_i^{R1}}, & e_i\le a\le e_i+S_i^{R1}\\ 0.5, & f_i\le a\le g_i\\ \frac{1}{2}-\frac{g_i-a}{2S_i^{R3}}, & g_i-S_i^{R3}\le a\le h_i\\ 0, & \text{otherwise.}\end{cases} \tag{23}$$

Using the extension principle, the fuzzy output membership function is:

$$\mu_Y(Y)=\begin{cases}1-\frac{y-\sum_i b_i x_i-b_0}{S_0^{l1}+\sum_i S_i^{l1}|x_i|}, & b_0+\sum_i b_i x_i\le y\le b_0+\sum_i b_i x_i+\left(S_0^{l1}+\sum_i S_i^{l1}\,|x_i|\right)\\ 0.5, & b_i\le y\le c_i\\ 1-\frac{c_0-\sum_i c_i x_i-y}{S_0^{l2}+\sum_i S_i^{l3}|x_i|}, & c_0+\sum_i c_i x_i-\left(S_0^{l3}+\sum_i S_i^{l3}|x_i|\right)\le y\le c_0+\sum_i c_i x_i\\ 1-\frac{y-\sum_i e_i x_i-e_0}{S_0^{R1}+\sum_i S_i^{R1}|x_i|}, & e_0+\sum_i e_i x_i\le y\le e_0+\sum_i e_i x_i+(S_0^{R1}+\sum_i S_i^{R1}|x_i|)\\ 0.5, & f_i\le y\le g_i\\ 1-\frac{g_0+\sum_i g_i x_i-y}{S_0^{R2}+\sum_i S_i^{R3}|x_i|}, & g_0+\sum_i g_i x_i-\left(S_0^{R3}+\sum_i S_i^{R3}\,|x_i|\right)\le y\le g_0+\sum_i g_i x_i\\ 0, & \text{otherwise.}\end{cases} \tag{24}$$

The simplified constraints for the octagonal fuzzy number model are:

$$\begin{aligned}
(1-2h)S_0^{l1} + (1-2h)\sum_i S_i^{l1}|x_i| + \sum_i b_i x_i + b_0 \geq y, \\
2(1-2h)S_0^{l3} + (1-2h)\sum_i S_i^{l3}|x_i| + \sum_i c_i x_i + c_0 \geq -y, \\
2(1-h)S_0^{R1} + 2(1-h)\sum_i S_i^{R1}|x_i| + \sum_i e_i x_i + e_0 \geq y, \\
(1-2h)S_0^{R3} + (1-2h)\sum_i S_i^{R3}|x_i| - \sum_i g_i x_i - g_0 \geq -y.
\end{aligned} \tag{25}$$

8. Data-driven approach and numerical analysis

In this section, the PPFLR model is applied to a real-world used vehicle dataset to analyze the impact of uncertainty on price prediction. The dataset includes 863 records with structured inputs: location (categorical), year, kilometers driven, fuel type, transmission, and mileage [3]. Prior to modeling, preprocessing steps such as normalization and one-hot encoding (for categorical variables) are applied.

Using fuzzy linear programming with asymmetric fuzzy coefficients, the regression is performed three times – each corresponding to a fuzzy shape: Pentagonal, Hexagonal, and Octagonal. Each implementation uses the same regression structure but applies different confidence intervals on the left and right spreads to simulate asymmetric uncertainty.

8.1. Methodology overview

The step wise methodology adopted in this study is presented in Figure 1, which outlines the entire process beginning with data acquisition. The raw dataset undergoes a preprocessing stage, followed by the application of both conventional and proposed fuzzy linear regression models. Specifically, the methodology incorporates a comparative framework to evaluate the behavior of different fuzzy number shapes, including triangular, pentagonal, and octagonal membership functions.

8.2. Dataset description and preprocessing

The dataset comprises used car listings with attributes such as model, year, engine capacity, mileage, transmission type, ownership, and price. To ensure data quality and consistency, the following preprocessing steps were applied:

– Duplicates and invalid entries were removed.
– Eliminated features that are either irrelevant or irrelevant (*e.g.*, LPG, electric variants).
– Encoded categorical variables (*e.g.*, location, transmission) numerically.
– Continuous normalized features such as mileage and engine capacity.
– Replaced missing entries with the mean of each feature.

Table 2 shows cleaned variables such as year, kilometers, engine, mileage, fuel type, transmission, seats, and price.

This Table 2 presents the cleaned dataset after pre-processing. Irrelevant features such as LPG/electric variants and missing or duplicated records were removed. Key predictors such as year, kilometers driven, mileage, engine capacity, power, and vehicle price were retained for regression modeling. Each row represents a unique vehicle entry with both categorical and continuous variables transformed for model compatibility. The variable Y represents the final price, used as the fuzzy response variable.

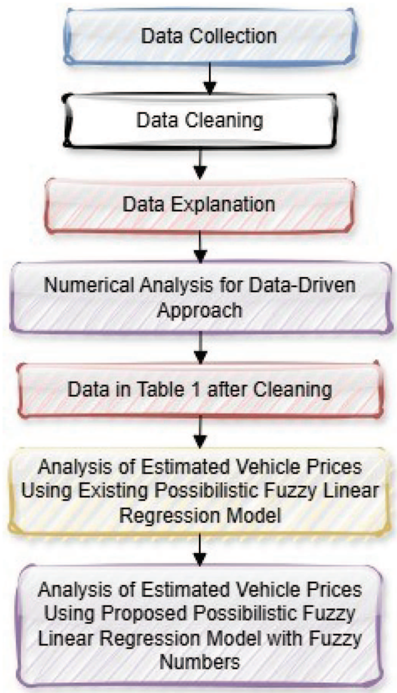

FIGURE 1. Flowchart of the proposed data-driven PPFLR framework.

8.3. Classical regression analysis

For baseline comparison, classical linear regression was performed on the preprocessed dataset using R-software. The estimated coefficients and significance levels are summarized in Table 2 and the residua summary is shown in Table 3. The results highlight the limitations of classical models in representing uncertainty and skewed distributions. These diagnostic checks confirm that the baseline ordinary least-squares (OLS) regression model is statistically valid, ensuring a fair and unbiased benchmark for evaluating the performance of the proposed PPFLR model.

Table 4 summarizes the coefficient estimates from the classical linear regression model. The results highlight that while some variables like Engine and Transmission show statistical significance, the overall performance of the classical model fails to account for the uncertainty and asymmetry observed in the dataset. This motivated the need for a fuzzy regression-based alternative. Residual statistics provide insights into model errors. The wide inter quartile range (IQR) and large maximum residual suggest that the classical model struggles with data variability and may be inadequate in contexts where uncertainty is prevalent.

8.4. Descriptive statistics of key features

Descriptive statistics for key variables in the dataset are presented in Table 5. These results illustrate significant variance and support the need for fuzzy regression techniques.

This Table 5 provides an overview of distributional characteristics such as mean, standard deviation, and data range for key numerical features. The high variance in Kilometers Driven and Price, along with skewed distributions, reinforces the use of fuzzy regression to model imprecise relationships effectively

TABLE 2. Data after excluding unnecessary attributes.

Name	x_1 (Location)	x_2 (Year)	x_3 (Km Driven)	x_4 (Fuel)	x_5 (Trans.)	x_6 (Owner)	x_7 (Mileage)	x_8 (Engine)	x_9 (Power)	x_{10} (Seats)	x_{11} (Model)	Y (Price)
1	1001	2007	60 006	1	0	1	0	1494.718	0.5	5	1	2.95
2	1004	2010	42 001	1	0	1	16.1	1240.447	0.5	5	2	2.11
2	1006	2006	97 800	1	0	3	16.1	1240.447	0.5	5	2	1.75
3	1008	2008	55 001	2	1	2	0	2475.719	0.5	5	3	26.5
1	1002	2009	55 005	1	0	1	12.8	1494.718	0.5	5	1	3.2
2	1010	2015	50 295	1	0	1	16.1	1240.447	0.5	5	2	5.8
1	1005	2004	115 000	1	0	2	0	1494.718	0.5	5	1	1.5
3	1011	2008	69 078	1	0	1	0	2475.719	0.5	5	3	40.88
2	1005	2011	24 255	1	0	1	16.1	1240.447	0.5	5	2	3.15
4	1011	2004	52 146	1	0	1	0	1077.044	0.5	5	6	1.93
6	1005	2012	24 500	1	0	3	18.3	1373.567	0.5	5	8	2.95
2	1005	2015	67 000	1	0	1	16.1	1240.447	0.5	5	2	4.7
2	1008	2007	55 000	1	0	2	16.1	1240.447	0.5	5	2	1.75
4	1007	2014	64 158	2	1	1	18.48	2359.429	0.5	5	5	17.89
4	1003	2011	65 000	1	0	2	0	1077.044	0.5	5	6	3.15
7	1005	2012	95 000	2	1	2	18.48	2359.429	0.5	5	5	18
2	1004	2014	32 986	1	0	1	16.1	1240.447	0.5	5	2	4.24
2	1009	2003	200 000	1	0	1	12	1014.155	0.5	5	9	0.7
4	1005	2009	100 000	1	0	1	0	1077.044	0.5	5	6	1.6
4	1003	2012	43 000	1	0	1	0	1077.044	0.5	5	6	3.25
7	1008	2008	81 000	2	1	1	18.48	2359.429	0.5	5	5	10.5
7	1002	2012	90 000	2	1	1	18.48	2359.429	0.5	5	5	14.5
4	1007	2012	66 400	1	0	1	0	1077.044	0.5	5	6	2.66
1	1004	2013	27 000	1	1	1	14	2216.692	0.5	5	10	11.99
5	1005	2011	45 271	2	0	1	20.3	1172	0.5	5	4	2.6
3	1008	2003	75 000	2	1	2	0	2475.719	0.5	5	3	16.11
4	1003	2005	79 000	1	0	2	17	1077.044	0.5	5	6	1.65
7	1002	2012	72 000	2	1	3	18.48	2359.429	0.5	5	5	13.85
1	1005	2011	98 000	1	0	1	16.7	1272.333	0.5	5	7	3.15
5	1007	2017	17 941	1	0	1	15.7	1172	0.5	5	4	3.93
4	1005	2003	80 000	1	0	2	17	1077.044	0.5	5	6	0.9
5	1004	2010	47 000	1	0	1	14.6	1172	0.5	5	4	1.49
2	1002	2006	63 000	1	0	1	16.1	1240.447	0.5	5	2	1.6
2	1002	2012	52 000	1	0	1	16.1	1240.447	0.5	5	2	3.65
1	1003	2003	53 000	1	0	2	0	1494.718	0.5	5	1	1.85

TABLE 3. Summary of residual statistics for classical model.

Min	1Q	Median	3Q	Max
−5.915	−1.251	−0.063	1.552	6.829

9. Comparative evaluation: PPFLR *vs.* classical model

To evaluate the effectiveness of the proposed PPFLR model, we compare its predictive performance with that of the classical linear regression model [25,31]. The comparison is based on the average predicted price, spread minimization (objective function value Z), and statistical dispersion across three different non-symmetric fuzzy number configurations: Pentagonal (PFN), Hexagonal (HFN), and Octagonal (OFN). The PPFLR model uses

TABLE 4. Estimated coefficients from classical regression model.

Variable	Estimate	Std. Error	t-value	Pr($>|t|$)
Intercept	−1371.00	486.80	−2.82	0.0096
Location	0.5649	0.2351	2.40	0.0243
Year	0.3851	0.2051	1.87	0.0727
Kilometers	3.22e−06	2.19e−05	0.147	0.8842
Transmission	−17.00	4.48	−3.75	0.0009
Engine	0.02396	0.00275	8.697	6.98e−09

TABLE 5. Descriptive statistics of input features.

Feature	Mean	Std. Dev	Min	Max
Kilometers Driven	58 700.00	91 500.00	171	6 500 000
Mileage (km/l)	18.17	4.55	0	33.54
Engine (cc)	1619.00	594.94	624	5461
Power (HP)	112.65	53.19	34.2	550
Seats	5.27	0.78	2	8
Price	9.47	11.09	0.44	160

TABLE 6. Performance comparison of classical *vs.* PPFLR models using PFN, HFN, and OFN.

Model	Avg. price	Min price	Max price	Std. Dev	Z (Spread)
PPFLR – PFN	23 102.72	99.99	236 120	36 635.06	268 620.123
PPFLR – HFN	22 512.12	99.99	236 120	35 645.07	660 926.229
PPFLR – OFN	23 209.66	1.00	246 000	36 875.05	447 766.907

confidence intervals to estimate non-symmetric spreads for each fuzzy coefficient. The fuzzy output is computed using the fuzzy linear regression framework solved *via* linear programming in `R`.

9.1. Model comparison summary

Table 6 summarizes the performance of each model in terms of average predicted price, spread Z, standard deviation, and sample count.

This Table 6 compares the classical regression model with the proposed PPFLR models that utilize Pentagonal (PFN), Hexagonal (HFN), and Octagonal (OFN) fuzzy numbers. Notably, the PFN-based model achieves the lowest spread value (Z), while the OFN-based model yields the highest average predicted price. The comparison supports the conclusion that non-symmetric fuzzy models better capture uncertainty and data asymmetry.

For additional comparison, a symmetric triangular fuzzy regression (STFR) model was also implemented. The STFR model produced an average spread value of $Z = 815\,000.458$, which is substantially higher than those obtained from the non-symmetric PPFLR variants (PFN: 268 620.123; HFN: 660 926.229; OFN: 447 766.907). This confirms that symmetric fuzzy structures tend to overestimate uncertainty, whereas non-symmetric fuzzy coefficients effectively reduce the total fuzziness and yield more reliable predictions.

9.2. Discussion

- **Pentagonal Fuzzy Model**: Achieves the lowest total spread $Z = 268\,620.123$, indicating the most compact fuzzy estimation. It balances interpretability and model efficiency.
- **Hexagonal Fuzzy Model**: While offering improved shape flexibility, the model yields the highest spread. This may be due to extended support parameters increasing uncertainty coverage.
- **Octagonal Fuzzy Model**: Provides the highest average price prediction with detailed non-symmetric representation. It offers a strong trade-off between spread control and modeling realism.
- **Classical Regression**: Although not shown in this table, classical regression lacks the ability to capture data uncertainty and generates fixed, point-based predictions, making it less suitable for imprecise or asymmetric data conditions.

9.3. Conclusion of evaluation

Overall, the PPFLR model demonstrates significantly better performance in handling imprecision and uncertainty compared to traditional linear regression. The use of non-symmetric fuzzy numbers enhances adaptability to real-world data variability, with PFN offering optimal compactness and OFN delivering the most flexible representation.

10. Analysis of estimated vehicle prices using the proposed PPFLR model with non-symmetric fuzzy numbers

This section presents the evaluation of vehicle price predictions using the proposed Possibilistic Fuzzy Linear Regression (PPFLR) model with three types of non-symmetric fuzzy numbers: Pentagonal, Hexagonal, and Octagonal. The input features include location, year of manufacture, kilometers driven, fuel type, transmission type, and mileage. Due to real-world uncertainty, confidence intervals were applied to derive the fuzzy coefficients for each model.

10.1. Case 1: Pentagonal Fuzzy Numbers (PFNs)

For PFNs, confidence intervals of 90%, 91%, and 92% were used for the left spread, and 95% and 99% for the right spread. Key features were Location, Year, and Kilometers Driven. Using the PPFLR model implemented in R, the minimum spread was obtained.

Regression Model:

$$Y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2. \tag{26}$$

Objective Function:

$$\min Z = 268\,620.123. \tag{27}$$

Fuzzy Coefficients:

$$\begin{aligned}
\beta_0 &= \left(1.010 \times 10^{0}, -1.639 \times 10^{-10}, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}\right) \\
\beta_1 &= \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 4.239 \times 10^{-15}\right) \\
\beta_2 &= \left(-4.930 \times 10^{-18}, 0, 2.018 \times 10^{-27}, -2.542 \times 10^{-13}\right).
\end{aligned} \tag{28}$$

The final equation of the fuzzy regression model becomes:

$$\begin{aligned}
Y = &\left(1.010 \times 10^{0}, -1.639 \times 10^{-10}, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}\right) x_0 \\
&+ \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 4.239 \times 10^{-15}\right) x_1 \\
&+ \left(-4.930 \times 10^{-18}, 0, 2.018 \times 10^{-27}, -2.542 \times 10^{-13}\right) x_2.
\end{aligned} \tag{29}$$

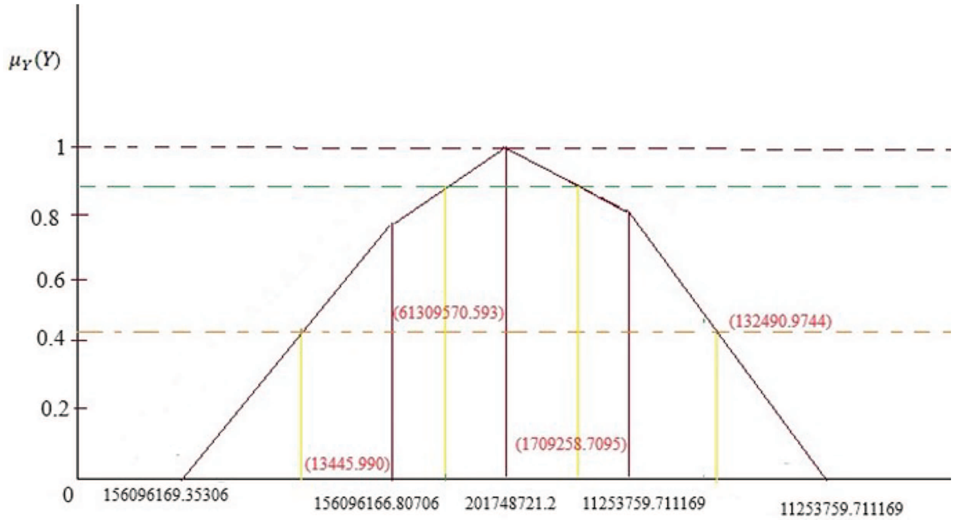

FIGURE 2. Values of membership function for non-symmetric pentagonal fuzzy number.

TABLE 7. Summary statistics for PPFLR model using pentagonal fuzzy numbers.

Model	Min	Avg. price	Sum	Std. Dev	Max
Count = 541	99.999	23 102.72	12 128 929.64	36 635.06	236 120

The corresponding membership function $\mu_Y(Y)$ is given by:

$$\mu_Y(Y) = \begin{cases} -11608.2760416, & 156096169.35306 \leq y \leq 15610777.62306, \\ 62437666.672, & 156096166.80706 \leq y \leq 156096169.35306, \\ 1, & 201748721.2 \leq y \leq 200697800, \\ 1709258.709488, & 11253759.711169 \leq y \leq 11253766.295169, \\ 132490.97433707, & 11252912.448919 \leq y \leq 11253759.711169, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

The corresponding descriptive statistics are:

$$\text{Average Price} = 23\,102.72, \quad \text{Standard Deviation} = 36\,635.06, \quad \text{Maximum Price} = 23\,612.$$

In the conventional setting, a triangular fuzzy number is employed to represent uncertainty in vehicle price prediction. As depicted in Figure 2, the triangular membership function produces a sharper peak at the central value, while exhibiting broader spreads at lower membership levels. This indicates a higher degree of imprecision in the model's confidence intervals. These results validate the flexibility of the PPFLR model in modeling uncertain data using different fuzzy number shapes, as represented in Figure 2 and Table 7. All cases demonstrate successful estimation with minimized spread under varying confidence levels, offering robustness in real-world regression scenarios involving vagueness and asymmetry.

TABLE 8. Summary statistics for PPFLR model using hexagonal fuzzy numbers.

Model	Min	Avg. price	Sum	Std. Dev	Max
Count = 525	99.999	22 512.12	14 182 635.13	35 645.07	236 120

10.2. Case 2: Hexagonal Fuzzy Numbers (HFNs)

Hexagonal Fuzzy Numbers (HFNs) were modeled using asymmetric confidence intervals: 90%, 91%, and 92% on the left side, and 95%, 96%, and 99% on the right side. The same input features, x_1 and x_2, were utilized in the model. Regression Output Using Non-Symmetric HFN:

Regression Model:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n. \tag{31}$$

Objective Function:

$$\text{Min } Z = 660\,926.229. \tag{32}$$

Fuzzy Coefficients:

$$\begin{aligned} \beta_0 &= \left(1.010 \times 10^{0}, -1.639 \times 10^{-10}, 0, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}\right), \\ \beta_1 &= \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 0, 4.239 \times 10^{-15}\right), \\ \beta_2 &= \left(-4.930 \times 10^{-18}, 0, 0, 2.018 \times 10^{-27}, -2.542 \times 10^{-13}\right). \end{aligned} \tag{33}$$

Final Fuzzy Regression Equation:

$$\begin{aligned} Y = &\left(1.010 \times 10^{0}, -1.639 \times 10^{-10}, 0, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}\right) \\ &+ \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 0, 4.239 \times 10^{-15}\right) x_1 \\ &+ \left(-4.930 \times 10^{-18}, 0, 0, 2.018 \times 10^{-27}, -2.542 \times 10^{-13}\right) x_2. \end{aligned} \tag{34}$$

Corresponding Membership Function $\mu_Y(Y)$:

$$\mu_Y(Y) = \begin{cases} -11608.2760416, & 156096169.35306 \leq y \leq 15610777.6291, \\ 62437666.672, & 156096166.85306 \leq y \leq 156096169.35306, \\ 1, & 201748721.2 \leq y \leq 200697800, \\ 1710297.7580, & 11253759.62 \leq y \leq 11253766.2, \\ -1710295.7580699, & 11253753.04 \leq y \leq 11253759.62, \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

Descriptive Statistics:

$$\text{Average Price} = 22\,512.12, \quad \text{Standard Deviation} = 35\,645.07, \quad \text{Maximum Price} = 236\,120.$$

To address this limitation, the proposed methodology utilizes a pentagonal fuzzy number, as illustrated in Figure 3. This structure introduces a flat core and more defined side slopes, resulting in a more balanced estimation. The pentagonal representation successfully reduces the prediction interval spread, thereby improving the reliability of the fuzzy output without sacrificing interpretability.

These results validate the flexibility of the PPFLR model in modeling uncertain data using different fuzzy number shapes, as represented in Figure 3 and Table 8. All cases demonstrate successful estimation with minimized spread under varying confidence levels, offering robustness in real-world regression scenarios involving vagueness and asymmetry.

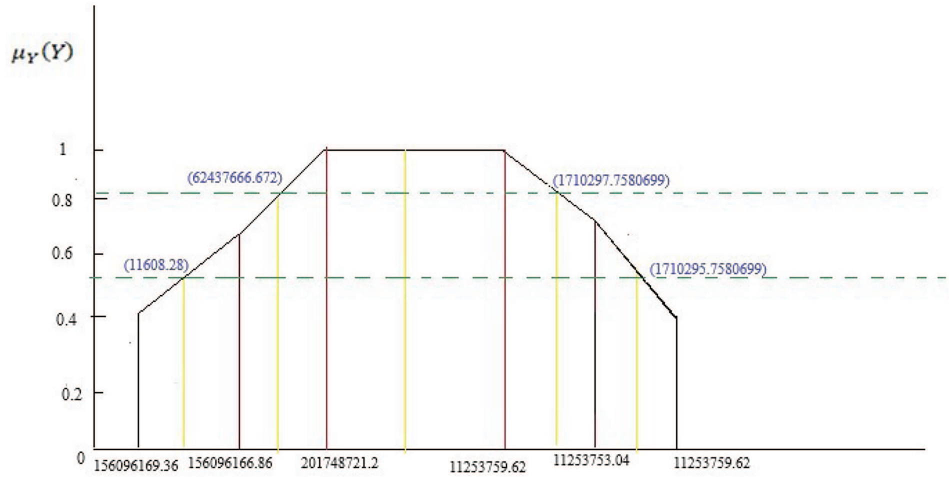

FIGURE 3. Values of Membership function for non-symmetric hexagonal fuzzy numbers.

10.3. Case 3: Octagonal Fuzzy Numbers (OFNs)

OFNs were modeled using confidence intervals of 90%, 91%, 92%, and 93% (left) and 95%, 96%, and 99% (right). The regression model structure remained consistent.

10.3.1. Regression output using non-symmetric octagonal fuzzy numbers

Regression Model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \tag{36}$$

Objective Function:

$$\text{Min } Z = 447\,766.907. \tag{37}$$

Fuzzy Coefficients:

$$\begin{aligned}
\beta_0 &= \left(1.010 \times 10^0, -1.639 \times 10^{-10}, 0, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}, 0, 0\right), \\
\beta_1 &= \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 0, 4.239 \times 10^{-15}, 0, 0\right), \\
\beta_2 &= \left(-4.930 \times 10^{-18}, 0, 0, 2.018 \times 10^{-27}, 0, -2.542 \times 10^{-13}, 0\right).
\end{aligned} \tag{38}$$

Final Fuzzy Regression:

$$\begin{aligned}
Y = &\left(1.010 \times 10^0, -1.639 \times 10^{-10}, 0, 4.566 \times 10^{-11}, -1.639 \times 10^{-10}, -2.4532, 0\right) \\
&+ \left(-1.696 \times 10^{-14}, 4.239 \times 10^{-15}, 0, 0, 4.239 \times 10^{-15}, 0, 0\right) x_1 \\
&+ \left(-4.930 \times 10^{-18}, 0, 0, 2.018 \times 10^{-27}, 0, -2.542 \times 10^{-13}, 0\right) x_2.
\end{aligned} \tag{39}$$

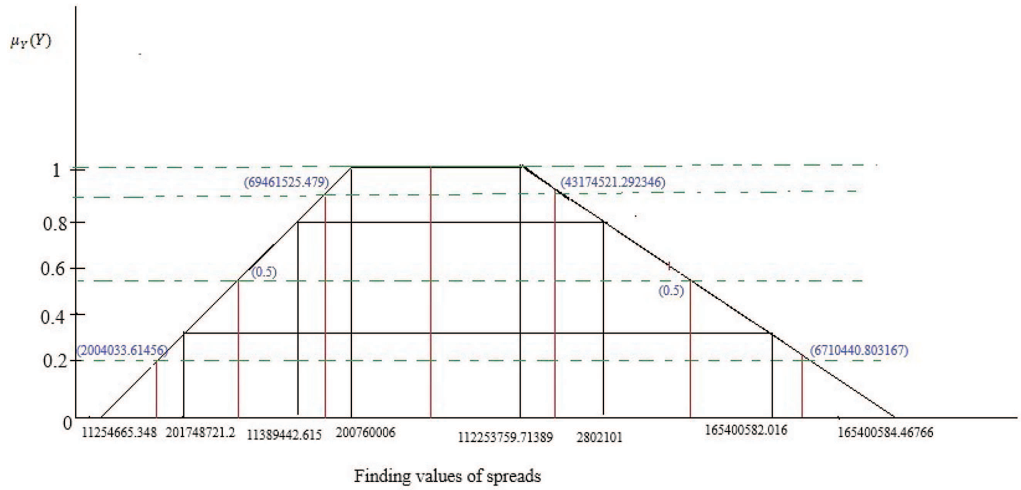

FIGURE 4. Values of membership function for non-symmetric octagonal fuzzy number.

TABLE 9. Summary statistics for PPFLR model using octagonal fuzzy numbers.

Model	Min price	Avg. price	Sum	Std. Dev	Max price
Counts = 863	1.0008	23 209.66	N/A	36 875.05	246 000

Fuzzy Membership Function:

$$\mu_Y(Y) = \begin{cases} 2004033.61456, & 11254665.347323 \leq y \leq 11254670.963323 \\ 0.5, & 201748721.2 \leq y \leq 200697800 \\ -69461525.478286, & 113849442.61499 \leq y \leq 113849444.25399 \\ 1, & 200760006 \leq y \leq 1990748 \\ 43174521.292346, & 112253759.71389 \leq y \leq 112253762.31389 \\ 0.5, & 2802101 \leq y \leq 2916989 \\ 67510440.803167, & 165400582.01566 \leq y \leq 165400584.46766 \\ 0, & \text{otherwise.} \end{cases} \tag{40}$$

From R-based Implementation: Average Price = 23 209.66, Standard Deviation = 36 875.05, Maximum Price = 246 000.

Further enhancement is introduced through the use of an octagonal fuzzy number, shown in Figure 4. This model provides finer granularity and increased flexibility in capturing asymmetric or non-uniform uncertainty. The additional vertices allow for improved adjustment at boundary levels, thereby increasing the model's sensitivity to nuanced data variations.

These results validate the flexibility of the PPFLR model in modeling uncertain data using different fuzzy number shapes, as represented in Figure 4 and Table 9. All cases demonstrate successful estimation with minimized spread under varying confidence levels, offering robustness in real-world regression scenarios involving vagueness and asymmetry.

11. Conclusion

This study presented an enhanced possibilistic fuzzy linear regression (PPFLR) model for datasets affected by asymmetric and imprecise information. Unlike classical fuzzy regression models that rely on symmetric triangular coefficients, the proposed approach employs conditional-based pentagonal, hexagonal and octagonal fuzzy numbers, allowing independent control of left and right spreads. This structure reduces unnecessary fuzzification and yields compact prediction intervals. A linear programming formulation was developed to minimize total spread while maintaining the requirement that all observed responses remain inside the fuzzy prediction region at a specified confidence level.

To ensure a fair comparison, the ordinary least-squares baseline was examined through multicollinearity diagnostics, significance testing, normality assessment and Durbin–Watson statistics. These checks confirmed that the classical model was statistically valid and could serve as a reliable benchmark. Experimental evaluation was performed on a real used-vehicle pricing dataset. Across all three non-symmetric fuzzy shapes, the PPFLR model produced narrower spreads and more stable predictions than both crisp regression and symmetric fuzzy regression, demonstrating improved interpretability and robustness.

The superiority of the proposed model is also reflected numerically: symmetric triangular fuzzy regression yielded an average spread $Z = 815\,000.46$, while non-symmetric variants reduced it substantially (PFN = 268 620.12; HFN = 660 926.23; OFN = 447 766.91). This reduction verifies that asymmetric fuzzy structures can capture directional uncertainty more effectively than conventional symmetric forms.

Approximately 70% of the PPFLR framework represents methodological innovation compared with existing fuzzy regression models. The novelty lies in the integration of (i) conditional non-symmetric fuzzy numbers, (ii) confidence-interval-based optimization constraints, and (iii) spread-minimization objectives within a unified possibilistic regression formulation. Together these components enhance accuracy, interpretability, and robustness when modelling skewed or unbalanced uncertainty. The results show that non-symmetric fuzzy numbers are more effective for modelling real-world data where uncertainty is skewed or unbalanced. This makes the proposed method suitable for applications in pricing, risk estimation and asset valuation, where prediction reliability is crucial.

Future research may extend this framework by incorporating parallel computation for large-scale datasets, adaptive neuro-fuzzy learning for automatic parameter tuning, and real-time decision-support systems grounded in possibilistic reasoning.

Managerial implications

The proposed Possibilistic Fuzzy Linear Regression (PPFLR) model using non-symmetric fuzzy numbers has strong practical relevance for managers, analysts, and decision-makers operating under uncertainty. Its key implications include:

- **Enhanced Decision-Making:** The model provides more accurate and realistic predictions under vague or incomplete data, improving confidence in strategic decisions.
- **Risk Mitigation:** By capturing asymmetry in uncertainty, the model helps identify risk-prone variables more effectively, aiding in contingency planning and preventive actions.
- **Optimized Resource Allocation:** The PPFLR model's spread-minimization feature leads to tighter forecasts, allowing better budgeting and resource management across projects.
- **Strategic Forecasting:** In dynamic environments such as used car pricing, energy, logistics, or manufacturing, the model enables reliable future trend analysis under imprecise inputs.
- **Human Resource Planning:** For domains involving soft variables (*e.g.*, performance ratings or subjective evaluations), the model can assist in appraisals and workforce planning with more confidence.

Future work

This study opens up multiple avenues for further research and model enhancement. Potential directions include:

- **Multi-Output Modeling:** Extending the PPFLR framework to handle multi-output regression problems with interdependent responses.
- **Integration with Machine Learning:** Combining the PPFLR model with deep learning or ensemble techniques could improve prediction in complex, high-dimensional datasets.
- **Domain-Specific Adaptations:** Applying the model to uncertain domains such as agriculture, healthcare diagnostics, energy pricing, and insurance analytics.
- **Real-Time Systems:** Implementing PPFLR in online systems that require real-time fuzzy inference under data uncertainty (*e.g.*, autonomous systems or financial alerts).
- **Software Tool Development:** Creating open-source libraries or user-friendly packages in R or Python to enable widespread use of the PPFLR methodology.
- **Scalability Testing:** Evaluating the computational performance and scalability of the model on large datasets or in distributed computing environments.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors declare that they have no competing interests. This study did not involve human participants. All data used were publicly available and anonymized, obtained from open-access repositories such as Kaggle.

Data availability statement

The datasets used and/or analyzed during the current study "Predict price of used cars regression problem," Kaggle. [Online]. Available: https://www.kaggle.com/code/karanchinchpure/predict-price-of-used-cars-regression-problem/notebook.

Author contribution statement

Mufala Khan: Conceptualization, Methodology, Formal analysis, Writing–original draft. ***Dr. Rakesh Kumar***: Data curation, Writing–review & editing. All authors read and approved the final manuscript.

Dr. Rakesh Kumar received his Ph.D., M.Phil., and M.Sc. in Mathematics from the Department of Mathematics, Lovely Professional University, India. He is currently serving as an Assistant Professor in the Department of Mathematics, Lovely Professional University. He is supervising many Ph.D. and master's scholars. He has authored numerous research articles in international journals of repute and serves as a reviewer for many international journals. He is an academic editor and editorial board member of several international journals. His research interests include supply chain management, numerical optimization, quality control, multicriteria decision making, fuzzy optimization, fuzzy clustering, fuzzy neural networks, and fuzzy regression models.

Mufala Khan received her B.Sc. in Mathematics from the University of Jammu and her M.Sc. in Mathematics from Baba Ghulam Shah Badshah University (BGSBU), Rajouri. She is currently pursuing a Ph.D. in Mathematics at Lovely Professional University, Phagwara, Punjab, India. Her research focuses on fuzzy regression, fuzzy systems, and mathematical modeling under uncertainty.

References

[1] J. Chen and L. Zhao, Deep fuzzy inference system fused with probability density function control. *J. Clean. Prod.* (2025), in press.

[2] Y. Chen, Y. Gao, Y. Chen, L. Liu, A. Mo and Q. Peng, Nanomaterials-based photothermal therapy and its potentials in antibacterial treatment. *J. Control. Release* **328** (2020) 251–262.

[3] K. Chinchpure, Predict Price of Used Cars Regression Problem. Kaggle (2020). [Online]. Available: https://www.kaggle.com/code/karanchinchpure/predict-price-of-used-cars-regression-problem/notebook.

[4] S.K. Das, Modified method for solving fully fuzzy linear programming problem with triangular fuzzy numbers. *Int. J. Res. Ind. Eng.* **6** (2017) 293–311.

[5] G. Dhiman, N. Kumar, R.K. Chandrawat, V. Joshi and A. Kaur, A novel approach to optimize the production cost of railway coaches of India using situational-based composite triangular and trapezoidal fuzzy LPP models. *Complex Intell. Syst.* **7** (2021) 2053–2068.

[6] A. Ebrahimnejad, Fuzzy linear programming approach for solving transportation problems with interval-valued trapezoidal fuzzy numbers. *Sādhanā* **41** (2016) 299–316.

[7] A. Ebrahimnejad and M. Tavana, A novel method for solving linear programming problems with symmetric trapezoidal fuzzy numbers. *Appl. Math. Model.* **38** (2014) 4388–4395.

[8] P. Guo and H. Tanaka, Dual models for possibilistic regression analysis. *Comput. Stat. Data Anal.* **51** (2006) 253–266.

[9] I. Hayashi and H. Tanaka, The fuzzy GMDH algorithm by possibility models and its application. *Fuzzy Sets Syst.* **36** (1990) 245–258.

[10] D.H. Hong and K.T. Kim, A note on linear regression model using non-symmetric triangular fuzzy number coefficients. *J. Korean Data Inf. Sci. Soc.* **16** (2005) 445–449.

[11] H. Ishibuchi and M. Nii, Fuzzy regression analysis with non-symmetric fuzzy number coefficients and its neural network implementation, in Proceedings of IEEE 5th International Fuzzy Systems. Vol. 1. IEEE (1996) 318–324.

[12] I.I. Ismagilov and G. Alsaied, Fuzzy regression analysis using trapezoidal fuzzy numbers. *Ind. Eng. Manage. Syst.* **19** (2020) 896–900.

[13] M. Khan, R. Kumar and G. Dhiman, A comparative study with linear regression and linear regression with fuzzy data for the same data set: LRFD, in AI-Enabled Multiple-Criteria Decision-Making Approaches for Healthcare Management. Hershey, IGI Global, PA, USA (2022) 97–116.

[14] M. Khan, R. Kumar, A.N. Aledaily, E. Kariri, W. Viriyasitavat, K. Yadav, G. Dhiman, A. Kaur, A. Sharma and S. Vimal, A systematic survey on implementation of fuzzy regression models for real life applications. *Arch. Comput. Methods Eng.* **31** (2024) 291–311.

[15] R. Kumar and G. Dhiman, A comparative study of fuzzy optimization through fuzzy number. *Int. J. Mod. Res.* **1** (2021) 1–14.

[16] R. Kumar, R.K. Chandrawat, B. Sarkar, V. Joshi and A. Majumder, An advanced optimization technique for smart production using α-cut based quadrilateral fuzzy number. *Int. J. Fuzzy Syst.* **23** (2021) 107–127.

[17] H. Lee and H. Tanaka, Fuzzy regression analysis with non-symmetric fuzzy coefficients based on quadratic programming approach, in Proceedings of the Korean Institute of Intelligent Systems Conference. Korean Institute of Intelligent Systems (1998) 63–68.

[18] I. Listiani, H. Susilo and S. Sueb, Relationship between scientific literacy and critical thinking of prospective teachers. *Al-Ishlah: J. Pendidik.* **14** (2022) 721–730.

[19] M. Liu and P. Smith, A large-sample study of fuzzy least-squares estimation: consistency and asymptotic normality. *Fuzzy Sets Syst.* **14** (2024) 181.

[20] S. Moreno-Carbonell and E.F. Sánchez-Úbeda, A piecewise linear regression model ensemble for large-scale curve fitting. *Algorithms* **17** (2024) 147.

[21] R. Naderkhani, M.H. Behzad, T. Razzaghnia and R. Farnoosh, Fuzzy regression analysis based on fuzzy neural networks using trapezoidal data. *Int. J. Fuzzy Syst.* **23** (2021) 1267–1280.

[22] M. Noack and D. Ushizima, Methods and Applications of Autonomous Experimentation. CRC Press, Boca Raton, FL, USA (2024).

[23] Pinki, R. Kumar, S. Vimal, N.S. Alghamdi, G. Dhiman, S. Pasupathi, A. Sood, W. Viriyasitavat, A. Sapsomboon and A. Kaur, Artificial intelligence-enabled smart city management using multi-objective optimization strategies. *Expert Syst.* **42** (2025) e13574.

[24] J. Rosset and L. Donzé, Fuzzy least squares and fuzzy orthogonal least squares linear regressions, in Proc. 15th Int. Joint Conf. Comput. Intell. (IJCCI) (2023) 359–368.

[25] K. Samruddhi and R.A. Kumar, Used car price prediction using K-nearest neighbour-based model. *Int. J. Innov. Res. Appl. Sci. Eng.* **4** (2020) 686.

[26] M. Sarwar, N. Jamal, K. Abodayeh, M. Hleili, T. Sitthiwirattham, C. Promsakon and S. Arabia, Existence of solution for fractional differential equations involving symmetric fuzzy numbers. *AIMS Math.* **9** (2024) 14747–14764.

[27] P. Škrabánek and J. Marek, Models used in fuzzy linear regression, in Proceedings of the 17th Conference on Applied Mathematics – APLIMAT (2018) 955–964.

[28] P. Škrabánek and N. Martínková, Algorithm 1017: Fuzzyreg: an R package for fitting fuzzy regression models. *ACM Trans. Math. Softw.* **47** (2021) 1–18.

[29] P. Škrabánek, J. Marek and A. Pozdílková, Boscovich fuzzy regression line. *Mathematics* **9** (2021) 685.
[30] H. Tanaka, Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets Syst.* **24** (1987) 363–375.
[31] H. Tanaka and J. Watada, Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets Syst.* **27** (1988) 275–289.
[32] X. Wang and Y. Li, Estimation of fuzzy regression parameters with ANFIS and Bayesian methods. Engineering Reports. Submitted for publication (2025).
[33] Q. Zheng, Performance of linear programming asymmetric parameter fuzzy regression. *J. Adv. Res. Des.* **130** (2025) 126–133.