STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–30.
Published online in International Academic Press (www.IAPress.org)

A Comparative Study of Full Z Number–Based Semi Parametric Regression Models: Accuracy, Structural Consistency, and Data Heterogeneity

Mahdi Ali Abdulhussein [1], Mehrdad Niaparast [1,*], Somayeh Ezadi [2]

[1]*Department of Statistics, Razi University, Iran*
[2]*Department of Mathematics, NT.C., Islamic Azad University, Tehran, Iran*

Abstract Simultaneous modeling of uncertainty and reliability in real-world data is one of the major challenges in regression analysis under uncertain conditions. Z-numbers, as a two-component framework, enable the concurrent representation of an imprecise value and the associated degree of confidence. However, developing Z-number-based regression models that possess adequate numerical accuracy, structural consistency, and the ability to cope with data heterogeneity remains challenging. In this study, for the first time, a fully semiparametric regression model based on Z-numbers is introduced, and three distinct solution methods are proposed. The first method, referred to as the baseline approach, provides a simple modeling framework by integrating the uncertainty and reliability components. The second method, referred to as the improved baseline method, explicitly separates the A and B components and applies inconsistency control mechanisms to preserve the conceptual coherence of Z-numbers. The third method, referred to as the clustered method, addresses data heterogeneity by proposing a Z-number-based clustering approach that enhances predictive accuracy through local model fitting. The results of the case study indicate that although the baseline model is structurally simpler and the improved baseline method provides satisfactory numerical stability, the clustered method exhibits the best performance in terms of error and similarity measures due to its ability to model data heterogeneity. Overall, the results demonstrate a trade-off among numerical accuracy, preservation of the conceptual structure of Z-numbers, and robustness against data heterogeneity; therefore, the proposed framework enables an informed selection of the appropriate method according to the application objectives and the nature of the data.

Keywords Z-Numbers, Semiparametric Regression, Uncertainty and Reliability, Structural Consistency, Data Heterogeneity, Z-Number-Based Clustering, Predictive Accuracy

DOI: 10.19139/soic-2310-5070-4456

1. Introduction

Semi-parametric regression, as one of the powerful tools in data analysis, enables the simultaneous modeling of parametric and nonparametric components [1] and therefore offers greater flexibility compared to fully parametric or nonparametric regression models [2]. This type of regression is particularly useful in problems where part of the relationship between explanatory variables and the response variable has a specified structure, while another part is unknown or dependent on time and environmental conditions. For this reason, semi-parametric regression has received considerable attention in fields such as economics [3], medicine [4], engineering [5], and social sciences [6].

To solve semi-parametric regression models, numerous classical methods have been proposed, including kernel estimators such as the Nadaraya–Watson method [7], penalized splines [8], series expansion–based approaches [9], and semiparametric least squares estimators [10]. In addition to these frequentist approaches, Bayesian

*Correspondence to: Mehrdad Niaparast (Email: niaparast@razi.ac.ir). Department of Statistics, Razi University, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

frameworks have also been developed to address more complex data structures, such as the Bayesian semi-parametric quantile regression approach for joint modeling of longitudinal responses proposed by Khazaei et al. [25]. Recent developments have also focused on improving estimation accuracy through shrinkage techniques, such as the linearized shrinkage estimator proposed by Rashad et al. [24]. Despite their effectiveness for precise and deterministic data, one of the fundamental limitations of these methods is their inability to handle data characterized by uncertainty, vagueness, and linguistic information a situation that inevitably arises in many real-world problems.

In response to this limitation, the concept of fuzzy logic was incorporated into the regression framework, leading to the development of fuzzy semiparametric regression models. In this approach, variables or model coefficients are considered as fuzzy numbers, and methods such as the α-cut [11], fuzzy programming [12], and fuzzy optimization techniques [13] are employed for parameter estimation. Although these models provide greater flexibility in handling uncertainty compared with classical methods, they often consider only the ambiguity component of the data and do not explicitly model the reliability or the degree of confidence associated with fuzzy information.

To overcome these shortcomings, the concept of Z-numbers was introduced [14], representing data as an ordered pair consisting of an uncertainty component (A) and a reliability component (B). This framework enables the simultaneous modeling of ambiguity and the degree of confidence in information and provides a more realistic representation of real-world data compared with classical fuzzy numbers [15, 16, 17, 18]. Compared to classical fuzzy sets, which capture only vagueness, Z-numbers offer a unique advantage by explicitly distinguishing between the uncertain value (A) and the reliability of that value (B). This dual structure provides a natural framework for representing linguistic information with confidence, making Z-numbers particularly suitable for real-world applications where data reliability varies across observations.

Nevertheless, Z-numbers introduce additional computational complexity due to their dual structure, and algebraic operations and distance measures are less straightforward than in classical fuzzy logic. Despite these challenges, the proposed semiparametric regression framework effectively addresses the computational difficulties through efficient optimization and penalty-based structural consistency, while leveraging the unique advantage of Z-numbers in modeling both uncertainty and reliability simultaneously. Despite the successful applications of Z-numbers in decision-making problems and intelligent systems [19], their use within the framework of semiparametric regression has not been systematically and comparatively investigated. In particular, a comprehensive framework for Z-number-based semiparametric regression that simultaneously considers numerical accuracy, structural consistency of the A and B components, and data heterogeneity has not been fully developed in the existing literature. This paper addresses this gap.

Only one study has been conducted on semiparametric regression based on Z-numbers [20], in which not all regression parameters are represented as Z-numbers, and a defuzzification stage is used to solve the model, which may lead to the loss of part of the informational structure of the data.

In this paper, for the first time, a comprehensive framework for fully Z-number-based semiparametric regression is proposed, which not only models uncertainty and reliability simultaneously but also systematically addresses the challenges of structural consistency and data heterogeneity. Unlike previous studies that have mainly focused on simplified or defuzzified versions, this research proposes three completely novel models within an evolutionary framework, each of which addresses a fundamental limitation in Z-number modeling.

First, a baseline semiparametric regression model based on Z-numbers is introduced, in which, for the first time, the uncertainty and reliability components are incorporated into the regression equation through a unified weighted structure. The main innovation of this model lies in providing a simple yet effective mechanism for transferring Z-number information into the regression space without completely reducing its fuzzy reliability structure; thus, this model can serve as a baseline for evaluating the numerical accuracy of subsequent developments.

In the second step, a generalized model with an explicit dual structure is proposed, in which the A (uncertainty) and B (reliability) components are estimated through separate sub-models while their interdependence is preserved via incompatibility penalties. The novelty of this approach lies in defining a regression framework that enables direct control of the internal inconsistency of Z-numbers and prevents the uncontrolled integration of these two components. For the first time, this model operationally incorporates the concept of structural consistency of Z-numbers into semiparametric regression and, from this perspective, represents a theoretical advancement over the baseline model.

Finally, in order to overcome the limitations of unified models when dealing with heterogeneous real-world data, a clustered semiparametric regression model based on Z-numbers is developed. The novelty of this method lies in the simultaneous integration of clustering concepts and Z-numbers within a semiparametric framework, such that each cluster is modeled as a homogeneous substructure with its own uncertainty and reliability characteristics. This approach, for the first time, enables local analysis of Z-number-based regression and plays an effective role in improving interpretability and managing data heterogeneity. Despite the remarkable progress in Z-number-based regression, the existing approaches exhibit several methodological limitations. Most available models are based on fully parametric formulations and are therefore unable to capture complex nonlinear relationships frequently encountered in real-world applications. Moreover, the treatment of the dual structure of Z-numbers, consisting of the uncertainty component (A) and the reliability component (B), differs considerably across existing studies. To better position the proposed framework within the current literature, Table 1 and Table 2 summarizes the main characteristics of representative Z-number-based regression models.

Table 1. Comparison of representative Z-number-based regression models

Method	**Regression Type**	**Fully Z-number Representation**	**Nonparametric Component**	**Structural Consistency (A–B)**	**Data Heterogeneity**
[17]	Linear	Partial	✗	✗	✗
[21]	Linear	✓	✗	Implicit	✗
[16]	Fuzzy regression	✓	✗	✓	✗
[20]	Semiparametric	Partial	✓	Limited	✗
Proposed Method 1	Fully Z-number	✓	Kernel	Penalty-based	✗
Proposed Method2	Fully Z-number	✓	Kernel	Explicitly	✗
Proposed Method 3	Fully Z-number	✓	Kernel +	Explicitly	✓

Table 2. Comparison of main limitations of representative Z-number-based regression models

Method	**Main Limitation**
[17]	Linear regression only; coefficients estimated using an improved neural network no explicit preservation of the Z-number structure.
[21]	Fully parametric model; unable to capture nonlinear relationships.
[16]	Provides a theoretical framework for Z-number fuzzy regression but does not consider semiparametric modeling or heterogeneous data.
[20]	Only part of the regression model is represented by Z-numbers, and the solution relies on a defuzzification stage, which may result in loss of information and does not explicitly preserve the dual Z-number structure.
Proposed Method 1	Global model; does not explicitly account does not explicitly account for heterogeneous data.
Proposed Method 2	Improved structural consistency but still assumes but still assumes a global model.
Proposed Method 3	Highest computational cost due to clustering and to clustering and local optimization

As can be seen from Table 1 and Table 2, previous studies mainly focus on parametric regression models and do not simultaneously address nonlinear modeling, explicit preservation of the dual Z-number structure, structural consistency between the uncertainty and reliability components, and heterogeneous data modeling. Although the recent semiparametric model of Abdulhussein et al. (2026) introduces nonparametric modeling, it still relies on a partially Z-number representation and a defuzzification stage during estimation. These observations reveal a clear research gap, motivating the development of a fully Z-number-based semiparametric regression framework that preserves both uncertainty and reliability throughout the modeling process while providing greater flexibility for complex nonlinear data. In summary, the innovation of this paper lies not in presenting a single model, but

in introducing a three-level family of Z-number-based semiparametric regression models that respectively aim to enhance numerical accuracy, structural coherence, and flexibility in handling data heterogeneity. Accordingly, the main research question is defined as follows:

> *In modeling Z-number-based semiparametric regression, how can an optimal balance be achieved among numerical accuracy, structural consistency of the Z-number components, and flexibility in handling data heterogeneity?*

The remainder of this paper is organized as follows. In Section 2, the fundamental concepts, including Z-numbers, fuzzy numbers, and the general framework of semiparametric regression, are introduced. Section 3 presents and explains the proposed methods for solving Z-number-based semiparametric regression. In Section 4, a case study along with a comparison of the numerical results of the three proposed methods is presented and analyzed. Finally, Section 5 summarizes the findings, addresses the research question, and outlines directions for future research, and the paper concludes with the list of references.

2. Fundamental Concepts

2.1. Definition of Z-numbers

Definition 2.1
Each Z-number is defined as an ordered pair $Z_i = (A_i, B_i)$, where A_i represents the fuzzy value component of the observation and B_i represents the confidence or reliability component associated with A_i [14].

In the Z-number literature, the terms "restriction" and "fuzzy value" are used interchangeably to refer to the A component, and "reliability" and "confidence" to refer to the B component.

2.2. Algebraic Operators of Z-Numbers

Definition 2.2 (Addition and Multiplication in Z-Number Algebra [26])
For two arbitrary Z-numbers $Z_1 = (A_1, B_1)$ and $Z_2 = (A_2, B_2)$, where

$$A_1 = (a_1^l, a_1^c, a_1^u), \quad B_1 = (b_1^l, b_1^c, b_1^u)$$
$$A_2 = (a_2^l, a_2^c, a_2^u), \quad B_2 = (b_2^l, b_2^c, b_2^u)$$

then

$$Z_1 \oplus_Z Z_2 = (A_1 \oplus_F A_2, B_1 \oplus B_2) \tag{1}$$

$$Z_1 \otimes_Z Z_2 = ((A_1 \otimes_F A_2), (B_1 \otimes B_2)) \tag{2}$$

where $\oplus_Z$ denotes the addition of two Z-numbers and $\otimes_Z$ denotes the multiplication of two Z-numbers. The operations $A_1 \oplus A_2$, $B_1 \oplus B_2$, $A_1 \otimes A_2$, and $B_1 \otimes B_2$ are respectively defined as follows:

$$A_1 \oplus_F A_2 = (a_1^l + a_2^l, a_1^c + a_2^c, a_1^u + a_2^u) \tag{3}$$

$$B_1 \oplus B_2 = (\min(b_1^l, b_2^l), \min(b_1^c, b_2^c), \min(b_1^u, b_2^u)) \tag{4}$$

$$A_1 \otimes_F A_2 = \left(\min(a_1^l a_2^l, a_1^l a_2^u, a_2^l a_1^u, a_1^u a_2^u),\ a_1^c a_2^c,\ \max(a_1^l a_2^l, a_1^l a_2^u, a_2^l a_1^u, a_1^u a_2^u)\right) \tag{5}$$

$$B_1 \otimes B_2 = (\min(b_1^l, b_2^l), \min(b_1^c, b_2^c), \min(b_1^u, b_2^u)) \tag{6}$$

2.3. Distance Measure Between Z-Numbers

Definition 2.3
For two arbitrary Z-numbers $Z_1 = (A_1, B_1)$ and $Z_2 = (A_2, B_2)$, where

$$A_1 = (a_1^l, a_1^c, a_1^u), \quad B_1 = (b_1^l, b_1^c, b_1^u)$$
$$A_2 = (a_2^l, a_2^c, a_2^u), \quad B_2 = (b_2^l, b_2^c, b_2^u)$$

the distance between the restriction components of a Z-number, denoted by $d_A(A_1, A_2)$, is defined as follows:

$$d_A(A_1, A_2) = \frac{|a_1^l - a_2^l| + 2|a_1^c - a_2^c| + |a_1^u - a_2^u|}{4} \tag{7}$$

and the distance between the reliability components of a Z-number, denoted by $d_B(B_1, B_2)$, is defined as:

$$d_B(B_1, B_2) = \frac{|b_1^l - b_2^l| + 2|b_1^c - b_2^c| + |b_1^u - b_2^u|}{4} \tag{8}$$

Definition 2.4
For an arbitrary Z-number $Z = (A, B)$, where $A = (a^l, a^c, a^u)$ and $B = (b^l, b^c, b^u)$ are triangular fuzzy numbers, we have:

$$\text{Width}(A) = a^u - a^l \tag{9}$$

$$\text{Centroid}(B) = C(B) = \frac{b^l + 2b^c + b^u}{4} \tag{10}$$

3. Fully Z-Number Semiparametric Regression Model and Its Solution Methods

In this subsection, a semipoparametric regression model is introduced entirely within the framework of Z-numbers, in which both the uncertainty of values and the degree of confidence associated with them are simultaneously incorporated into the modeling process. Unlike the model presented in [20], where only part of the model structure is defined based on Z-numbers, in the proposed model all components of the model, including the regression coefficients and the nonparametric component, are expressed in the form of Z-numbers. This approach extends classical and fuzzy regression models while enabling the simultaneous modeling of quantitative uncertainty and reliability uncertainty, making it particularly suitable for inherently uncertain data, including medical data.

3.1. Fully Z-Number Semiparametric Regression Model

The fully Z-number semiparametric regression model is defined as follows:

$$[y_i]_Z = [\beta]_Z \otimes_Z [x_i]_Z \oplus_Z [f(t_i)]_Z \tag{11}$$

where $[y_i]_Z = (y_{Ai}, y_{Bi})$ is the response variable expressed as a Z-number, $[x_i]_Z = (x_{Ai}, x_{Bi})$ is the explanatory variable expressed as a Z-number, $[\beta]_Z = (A_\beta, B_\beta)$ is the regression parameter based on Z-numbers, $[f(t_i)]_Z = (f_A(t_i), f_B(t_i))$ is the time-dependent nonparametric component expressed as a Z-number, and t_i is the normalized time variable. The operators $\otimes_Z$ and $\oplus_Z$ denote the multiplication and addition operations in Z-number algebra, respectively.

In this paper, the normalized time variable for the i-th observation is defined as

$$t_i = \frac{i-1}{n-1}, \quad i = 1, \ldots, n, \tag{12}$$

such that $t_i \in [0,1]$. This definition maps the temporal index of observations onto the unit interval and enables stable and scale-independent modeling of the nonparametric component $f(t)$.

By decomposing the Z-number components, for the model $[y_i]_Z = (y_{Ai}, y_{Bi})$ in equation (11), and according to [21], we obtain:

$$y_{A_i} = (\beta_A \otimes x_{A_i}) \oplus f_A(t_i) \tag{13}$$

$$y_{B_i} = (\beta_B \otimes x_{B_i}) \oplus f_B(t_i) \tag{14}$$

In the following, we present three methods for solving the Z-number-based semiparametric regression model.

3.2. Solution of the Fully Z-Number Semiparametric Regression Model (Baseline Method)

Consider the Z-number-based semiparametric regression model defined in equation (11) as follows:

$$[y_i]_Z = [\beta]_Z \otimes_Z [x_i]_Z \oplus_Z [f(t_i)]_Z$$

Assume that for each observation $i = 1, \ldots, n$, the input variable $[x_i]_Z = (x_{Ai}, x_{Bi})$ and the output variable $[y_i]_Z = (y_{Ai}, y_{Bi})$ are defined as Z-numbers, such that x_{A_i}, y_{A_i}, x_{B_i}, and y_{B_i} are triangular fuzzy numbers defined as

$$x_{A_i} = (x^l_{A_i}, x^c_{A_i}, x^u_{A_i}), \quad x_{B_i} = (x^l_{B_i}, x^c_{B_i}, x^u_{B_i}), \tag{15}$$

$$y_{A_i} = (y^l_{A_i}, y^c_{A_i}, y^u_{A_i}), \quad y_{B_i} = (y^l_{B_i}, y^c_{B_i}, y^u_{B_i}) \tag{16}$$

Furthermore, for $[\beta]_Z = (\beta_A, \beta_B)$, we assume

$$\beta_A = (\beta^l_a, \beta^c_a, \beta^u_a), \quad \beta_B = (\beta^l_b, \beta^c_b, \beta^u_b) \tag{17}$$

and for $[f(t_i)]_Z = (f_A(t_i), f_B(t_i))$, we assume

$$f_A(t_i) = (f^l_a, f^c_a, f^u_a), \quad f_B(t_i) = (f^l_b, f^c_b, f^u_b) \tag{18}$$

According to relations (13)–(16), we have

$$y_{A_i} = (\beta_A \otimes x_{A_i}) \oplus f_A(t_i) = (y^l_{A_i}, y^c_{A_i}, y^u_{A_i}) \tag{19}$$

$$y_{B_i} = (\beta_B \otimes x_{B_i}) \otimes f_B(t_i) = (y^l_{B_i}, y^c_{B_i}, y^u_{B_i}) \tag{20}$$

where

$$\beta_A \otimes x_A = \big(\min(\beta^l_a x^l_{A_i}, \beta^l_a x^u_{A_i}, \beta^u_a x^l_{A_i}, \beta^u_a x^u_{A_i}),\ \beta^c_a x^c_{A_i},\ \max(\beta^l_a x^l_{A_i}, \beta^l_a x^u_{A_i}, \beta^u_a x^l_{A_i}, \beta^u_a x^u_{A_i})\big) \tag{21}$$

$$\beta_B \otimes x_B = \big(\min(\beta^l_b, x^l_{B_i}), \min(\beta^c_b, x^c_{B_i}), \min(\beta^u_b, x^u_{B_i})\big) \tag{22}$$

For each component $k \in \{l, c, u\}$, the value $f^k_A(t_i)$ in equation (18) is estimated using the Nadaraya–Watson kernel regression as follows [22]:

$$f^k_A(t_i) = \sum_{j=1}^{n} \omega^{(k)}_{ij}(t_i)\big(y^k_{A_j} \ominus \beta^k_{A_j} x^k_{A_j}\big) \tag{23}$$

where the weights are defined as

$$\omega_{ij}^{(k)}(t_i) = \frac{K((t_j - t_i)/h_k)}{\sum_{l=1}^{n} K((t_l - t_i)/h_k)} \tag{24}$$

In this study, the triweight kernel is employed, which is defined as

$$K(u) = \begin{cases} \frac{15}{16}(1-u^2)^3, & |u| \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

To model the temporal variation of the reliability component $f_B^k(t)$ in equation (18), a parametric logistic function is used:

$$f_B^k(t) = \frac{1}{1+\exp(-\theta_k t)}, \quad k \in \{l, c, u\} \tag{26}$$

where θ_k are unknown parameters that must be estimated during the model estimation process (typically using an objective function and optimization methods such as least squares or fuzzy loss functions) for each component $k \in \{l, c, u\}$. Here, in order to ensure numerical stability, the values of $f_B^k(t)$ are restricted to the interval $[0.1, 0.9]$.

Now, to solve model (11), the unknown parameters namely $\beta_Z = (\beta_A, \beta_B)$, the bandwidth parameters h_k (for f_A), and the slope parameters θ_k (for f_B) must be determined such that the total error between $[y_i]_Z$ and $[\beta]_Z \otimes_Z [x_i]_Z \oplus_Z [f(t_i)]_Z$ is minimized.

To this end, the model parameters are estimated by solving the following optimization problem:

$$\min_{\beta_Z, \theta, h} L(\beta_Z, \theta, h) \tag{27}$$

where $L(\beta_Z, \theta, h)$ is the objective function, which simultaneously ensures goodness of fit to the data, internal consistency of the Z-numbers, and structural stability of the model. It is defined as follows:

$$L = L_{\text{fit}} + P_{\text{inc}} + R \tag{28}$$

where

$$L_{\text{fit}} = \frac{1}{n}\sum_{i=1}^{n}\left(w_A \frac{d_A(y_{iA}, \hat{y}_{iA})}{\text{width}(y_{iA})} + w_B d_B(y_{iB}, \hat{y}_{iB})\right) \tag{29}$$

Here, w_A and w_B (the weights associated with the fitting errors of components A and B) are considered known hyperparameters. The quantities $d_A(\hat{y}_{iA}, y_{iA})$ and $d_B(\hat{y}_{iB}, y_{iB})$ are calculated according to formulas (7) and (8), respectively.

P_{inc} is the incompatibility penalty defined as follows:

$$\begin{aligned} P_{\text{inc}} = {} & \lambda_\beta \big(\text{width}(\beta_A) - k_\beta(1 - \text{centroid}(\beta_B))\big)^2 \\ & + \lambda_f \sum_i \big(\text{width}(f_A(t_i)) - k_f(1 - \text{centroid}(f_B(t_i)))\big)^2 \\ & + \lambda_y \sum_i \big(\text{width}(\hat{y}_{A_i}) - k_y(1 - \text{centroid}(\hat{y}_{B_i}))\big)^2 \end{aligned} \tag{30}$$

where k_β, k_f, and k_y are scaling parameters that control the relationship between the degree of fuzziness of the value component and the reliability component in a Z-number. In this paper, for simplicity and without loss of generality, it is assumed that $k_\beta = k_f = k_y = 1$.

The functions width$(\cdot)$ and centroid$(\cdot)$ are obtained according to relations (9) and (10), respectively.

he squared form of the incompatibility penalty in Equation (30) was chosen for two primary reasons. First, it provides a smoother gradient specifically, a linear gradient $2x$ which facilitates more stable and consistent updates during the L-BFGS-B optimization process, compared to the discontinuous gradient of an absolute value penalty.

Second, the squared penalty disproportionately penalizes large deviations from the ideal Z-number consistency relationship, thereby encouraging the model to maintain smaller incompatibilities throughout the estimation process. This design choice is specific to the baseline method; in the improved method (Method 2, Equations 38–40), the penalty is reformulated as a first-order absolute difference to provide more explicit interpretability of the incompatibility magnitude.

In relation (28), R is the regularization term, which is added to the objective function in order to control the model complexity and ensure the smoothness of the Z-number components. It is computed as follows:

$$R = \mu_A \|\beta_A\|_2^2 + \mu_B \|\beta_B\|_2^2 + \mu_t \sum_{i=2}^{n} \|f_B(t_i) - f_B(t_{i-1})\|_2^2 \tag{31}$$

where μ_A, μ_B, and μ_t are regularization parameters that control, respectively, the shape of the fuzzy coefficients, the smoothing of the reliability parameters, and the temporal smoothing of the nonparametric component.

In the implementation of the model, the L_2-norm for the fuzzy coefficient $\beta_A = (l, c, u)$ is computed in a way that explicitly controls the width and symmetry of the triangular fuzzy number:

$$\|\beta_A\|_2^2 = (\beta_a^u - \beta_a^l)^2 + ((\beta_a^u - \beta_a^c) - (\beta_a^c - \beta_a^l))^2.$$

For the reliability parameters, we define

$$\|\beta_B\|_2^2 = (\beta_b^l)^2 + (\beta_b^c)^2 + (\beta_b^u)^2.$$

The temporal smoothing of the reliability component is also imposed as

$$\sum_{i=2}^{n} \big(f_B(t_i) - f_B(t_{i-1})\big)^2.$$

Problem (27) is solved using the L-BFGS-B algorithm [23], with appropriate bounds imposed on the parameters. In this method, the parameter vector

$$\theta = (\beta_a^l, \beta_a^c, \beta_a^u, \beta_b^l, \beta_b^c, \beta_b^u, \theta_1, \theta_2, \theta_3, h_l, h_c, h_u)$$

is optimized under suitable parameter bounds. The L-BFGS-B algorithm is selected due to its high efficiency in solving nonlinear optimization problems with box constraints. Specifically, θ_1, θ_2, and θ_3 represent the slope coefficients of the logistic functions governing the temporal evolution of the left, center, and right components of the reliability function $f_B(t)$. Furthermore, the parameters h_l, h_c, and h_u serve as the bandwidths of the Nadaraya–Watson kernel estimators used to obtain the nonparametric components $f_A^l(t), f_A^c(t)$, and $f_A^u(t)$. These bandwidths control the degree of smoothing applied to each part of the fuzzy valued function, allowing for flexible modeling of potential asymmetry in the underlying fuzzy structure. The proposed model inherently controls the inconsistency between uncertainty and reliability and enables the simultaneous modeling of fuzzy structure, reliability, and temporal dynamics, a feature that has not been considered in many existing models based on fuzzy numbers or Z numbers.

3.3. Second Proposed Method for Solving the Z-Number Based Semiparametric Regression (Improved Baseline Method)

Consider the Z-number-based semiparametric regression model defined in equation (11):

$$[y_i]_Z = \beta_Z \otimes_Z [x_i]_Z \oplus_Z [f(t_i)]_Z$$

For each component $k \in \{l, c, u\}$, the value $f_A^k(t_i)$ of $[f(t_i)]_Z = (f_A(t_i), f_B(t_i))$ appearing in equation (11) is estimated using equations (23)–(25).

For modeling the time-dependent reliability component $f_B^k(t_i)$, a smooth parametric structure based on sinusoidal functions is considered:

$$\begin{cases} f_B^l(t) = \alpha_l + \gamma_l \sin(2\pi t + \phi_l)\theta_1, \\ f_B^c(t) = \alpha_m + \gamma_m \sin(2\pi t + \phi_m)\theta_2, \\ f_B^u(t) = \alpha_u + \gamma_u \sin(2\pi t + \phi_u)\theta_3, \end{cases} \tag{32}$$

where $\alpha_\bullet$ represents the baseline level of reliability, $\gamma_\bullet > 0$ denotes the temporal oscillation amplitude, $\phi_\bullet$ is the phase shift, $\theta_1, \theta_2, \theta_3 \in [0, 1]$ are coefficients controlling the oscillation intensity, and $t \in [0, 1]$ is the normalized time variable. The sinusoidal function is employed to model the temporal evolution of the reliability component, as it provides a bounded and smooth parametric structure capable of capturing potential non-monotonic or cyclical variations in confidence levels over time. This specification offers greater flexibility than the logistic function used in the baseline model, particularly in applications where reliability may exhibit periodic fluctuations due to measurement conditions or environmental factors.

The model parameters are estimated by solving the following optimization problem:

$$\min_{\beta_Z, \theta, h_L, h_C, h_R} L(\beta, \theta, h) \tag{33}$$

In equation (33), the objective function $L(\beta, \theta, h)$ simultaneously ensures data fitting accuracy, internal consistency of Z-numbers, and structural stability of the model, and is defined as

$$L(\theta) = L_{\text{fit}} + L_{\text{incompat}} + L_{\text{reg}} \tag{34}$$

where L_{fit} denotes the data fitting error, computed as

$$L_{\text{fit}} = \sum_{i=1}^{n} \left(w_A d_A(y_{A_i}, \hat{y}_{A_i}) + w_B d_B(y_{B_i}, \hat{y}_{B_i}) \right) \tag{35}$$

where ω_A and ω_B (the weights associated with the fitting errors of components A and B) are treated as known hyperparameters. The quantities $d_A(\cdot)$ and $d_B(\cdot)$ represent the distances between triangular fuzzy numbers, obtained using equations (7) and (8), respectively.

The regularization term L_{reg}, added in equation (34), is formulated to control the model complexity and prevent the overfitting phenomenon. It is defined as

$$L_{\text{reg}} = \mu_A \|\beta_A\|_2^2 + \mu_B \|\beta_B\|_2^2 + \mu_t \sum_{i=2}^{n} \|f_B(t_i) - f_B(t_{i-1})\|^2 \tag{36}$$

In this equation, the regularization hyperparameters are interpreted as follows:

- μ_A and μ_B: These coefficients are used as L_2-penalties for the fuzzy regression coefficients β_A and β_B, respectively. Their purpose is to control the magnitude of these coefficients in order to maintain estimation stability.
- μ_t: This coefficient represents the smoothness penalty that targets the temporal dependency. By computing the sum of squared differences between consecutive values of f_B over time ($\sum_i \|f_B(t_i) - f_B(t_{i-1})\|^2$), this term ensures that the changes in the time-dependent reliability component f_B between successive data points are not excessively abrupt or unstable. This regularization directly influences the behavior of the newly introduced sinusoidal model of f_B.

In equation (34), L_{incompat} is defined as the sum of three independent incompatibility penalties, whose purpose is to enforce the fundamental inverse principle in Z-numbers, namely $\text{Width}(Z) \approx k(1 - \text{Centroid}(Z))$ at different levels of the model. This term is defined as

$$L_{\text{incompat}} = P_\beta + P_f + P_y \tag{37}$$

where P_β represents the incompatibility penalty associated with the regression parameters, defined as

$$P_\beta = \lambda_\beta \left| \text{Width}(\beta_A) - k_\beta(1 - \text{centroid}(\beta_B)) \right| \tag{38}$$

This means that if β_A has a large width while β_B indicates high reliability, the model is penalized. Consequently, the regression coefficients represented as Z-numbers remain logically consistent.

On the other hand, P_f is the incompatibility penalty for the time-dependent nonparametric component, defined as

$$P_f = \lambda_f \sum_i \left| \text{Width}(f_A(t_i)) - k_f(1 - \text{centroid}(f_B(t_i))) \right| \tag{39}$$

That is, at each time point t_i, if the uncertainty level $f_A(t_i)$ is large while the corresponding reliability $f_B(t_i)$ remains high, the model is penalized. This term precisely reflects the role of Z-numbers in the nonparametric part of the model.

Finally, P_y denotes the incompatibility penalty of the predicted output, and is defined as

$$P_y = \lambda_y \sum_i \left| \text{Width}(\hat{y}_i) - k_y(1 - \text{centroid}(\hat{y}_i)) \right| \tag{40}$$

meaning that the final model output must also constitute a valid Z-number, not only the parameters and intermediate components.

Here, Width$(\cdot)$ and Centroid$(\cdot)$ are obtained according to equations (9) and (10), respectively. The penalty hyperparameters consist of the weight coefficients $(\lambda_\beta, \lambda_f, \lambda_y)$, which control the strength of each incompatibility penalty, and the scaling constants (k_β, k_f, k_y), which are used to align the measures of uncertainty and reliability in all three penalty terms. These six parameters, together with the regularization coefficients μ, are regarded as known hyperparameters of the model, determined prior to optimization.

3.4. Third Method: An Integrated Framework for Cluster-Based Semiparametric Regression with Z-Numbers

In this section, an integrated framework for cluster-based semiparametric regression with Z-numbers is presented. The primary goal of this proposed method is to model the input–output relationships under uncertainty while simultaneously preserving both uncertainty and reliability information, without resorting to any reduction to crisp values or the α-cut method.

To proceed with this method, it is first necessary to introduce the required concepts.

Definition 3.1
For an arbitrary Z-number $Z = (A, B)$, where $A = (a^l, a^c, a^u)$ and $B = (b^l, b^c, b^u)$ are triangular fuzzy numbers, the structural reliability distance is defined as

$$d_{\text{rel}} = \left| \frac{\text{Width}(y_{Ai})}{\text{Centroid}(y_{Bi}) + \varepsilon} - \frac{\text{Width}(\hat{y}_{Ai})}{\text{Centroid}(\hat{y}_{Bi}) + \varepsilon} \right| \tag{41}$$

where Width $y_{Ai}, \hat{y}_{Ai}$ and Centroid $y_{Bi}, \hat{y}_{Bi}$ are computed using formulas (9) and (10), respectively.
To ensure numerical stability and prevent division by zero or excessively small values, a sufficiently small positive constant $\varepsilon > 0$ is added to the denominator in both terms of the structural reliability distance. This is a standard practice in numerical optimization to avoid overflow and maintain stable computations. The value of is chosen small enough to have a negligible effect on the computed distances while effectively preventing numerical instabilities.
This measure directly models the dependency between the width of uncertainty and the level of reliability.

Definition 3.2
For two input–output samples represented by Z-numbers

$$Z_i = \big((x_A^i, x_B^i), (y_A^i, y_B^i)\big), \tag{42}$$

where x_A^i, y_A^i denote the centroid components, and x_B^i, y_B^i denote the width components for sample i, the Integrated Distance between two samples Z_i and Z_j is defined as

$$D(Z_i, Z_j) = \alpha D_A + \beta D_B + \gamma d_{\text{rel}}, \quad \alpha + \beta + \gamma = 1 \tag{43}$$

where α, β, γ are scaling weights corresponding to the centroid component, width component, and the structural/incompatibility measure, respectively.

The components D_A and D_B are defined as

$$D_A = w_1 d_A(x_A^i, x_A^j) + w_2 d_A(y_A^i, y_A^j), \tag{44}$$

$$D_B = w_3 d_B(x_B^i, x_B^j) + w_4 d_B(y_B^i, y_B^j). \tag{45}$$

3.4.1. Integrated Z-Number Fuzzy Clustering In this paper, data clustering is performed using the Fuzzy C-Means (FCM) algorithm. In this approach, the distance between each data sample Z_i and the cluster center V_k (which is itself a Z-number) is directly measured using the Integrated Distance $D(Z_i, V_k)$ defined in equation (43). The objective function of FCM is defined as

$$J = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m \big(D(Z_i, V_k)\big)^2, \tag{46}$$

where m is the fuzzification parameter. V_k denotes the center of cluster k (represented as a Z-number) and is computed as

$$V_k = \frac{\sum_{i=1}^{n} u_{ik}^m Z_i}{\sum_{i=1}^{n} u_{ik}^m} \tag{47}$$

Here, u_{ik} represents the membership degree of sample i in cluster k. In the Fuzzy C-Means algorithm, when the cluster centers V_k are fixed, the membership degree u_{ik} is calculated as

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D(Z_i, V_k)}{D(Z_i, V_j)} \right)^{2/(m-1)}}, \tag{48}$$

where $D(Z_i, V_k)$ denotes the Integrated Distance, computed using equation (43).

3.4.2. Solving the Z-Number Based Semiparametric Regression (Method 3) Consider the Z-number-based semiparametric regression model defined in equation (11) as

$$[y_i]_Z = \beta_Z \otimes_Z [x_i]_Z \oplus_Z [f(t_i)]_Z$$

For each component $k \in \{L, C, R\}$, the value $f_A^k(t_i)$ from $[f(t_i)]_Z = (f_A(t_i), f_B(t_i))$ in equation (11) is estimated using relations (23)–(25).

The reliability component is modeled as a parametric logistic function as follows:

$$f_B(t) = a + \frac{b}{1 + \exp(-\theta(t - 0.5))} \tag{49}$$

To ensure numerical stability, the values of $f_B^k(t)$ are constrained within the interval $[0.3, 0.95]$. Based on the clustering performed above, the parameters β_Z and $f(t)$ are locally optimized for each cluster. Consequently, the index sets I_β and I_f consist of samples belonging to a specific cluster, allowing the local structure of the data to be incorporated into the model.

The final objective function in this method is defined as

$$L = L_{\text{fit}} + L_{\text{inc}} + L_{\text{reg}} + L_{\text{struct}} \tag{50}$$

where L_{fit} denotes the fitting error, computed as

$$L_{\text{fit}} = \sum_{i=1}^{n} \big(w_A d_A(Y_i, \hat{Y}_i) + w_B d_B(Y_i, \hat{Y}_i)\big) \tag{51}$$

On the other hand, L_{reg} represents the regularization terms, defined as

$$L_{\text{reg}} = \mu_A \|\beta_A\|^2 + \mu_B \|\beta_B\|^2 + \mu_t \sum_{i=2}^{n} \|f_B(t_i) - f_B(t_{i-1})\|^2 \tag{52}$$

In equation (50), L_{inc} represents the incompatibility term. This term ensures that the predictive model properly incorporates the uncertainty and reliability information of both inputs and outputs during modeling. It is typically defined based on the structural reliability distance (d_{rel}) or the Integrated Distance (D).

$$L_{\text{inc}} = \lambda \sum_{i=1}^{n} \big| d_{\text{rel}}(Y_i, \hat{Y}_i) - d_{\text{target}} \big| \tag{53}$$

where $d_{\text{rel}}(Y_i, \hat{Y}_i)$ is computed using equation (41) based on the actual output Y_i and the predicted output $\hat{Y}_i$. Here, λ is a hyperparameter controlling the weight of the incompatibility term, and d_{target} is a reference value (or zero) that specifies the desired level for preserving the uncertainty ratio between the outputs.

In equation (50), L_{struct} denotes the structural regularization term, which is introduced to ensure that the internal relationship of Z-numbers is preserved in the predicted outputs:

$$\begin{aligned} L_{\text{struct}} = & \sum_{i \in I_\beta} \big| \text{Width}(\beta^A_{Z_i}) - k(1 - \text{Centroid}(\beta^B_{Z_i})) \big| \\ & + \sum_{i \in I_f} \big| \text{Width}(f_{A_i}(t)) - k(1 - \text{Centroid}(f_{B_i}(t))) \big| \\ & + \sum_{i \in I_{\hat{Y}}} \big| \text{Width}(\hat{Y}_{A_i}) - k(1 - \text{Centroid}(\hat{Y}_{B_i})) \big| \end{aligned} \tag{54}$$

where λ_{struct} is a control hyperparameter that determines the weight of this structural constraint.

3.5. Model Evaluation

To assess the performance of the proposed models, the prediction error is computed separately for the fuzzy value component and the reliability component.

3.5.1. Fuzzy Component Error The Mean Absolute Error (MAE) for the fuzzy part is defined as:

$$\text{MAE}_A = \frac{1}{n} \sum_{i=1}^{n} d_A(y_{iA}, \hat{y}_{iA}), \tag{55}$$

where $d_A(\cdot, \cdot)$ denotes the fuzzy distance between the actual and the predicted values of the fuzzy component, calculated according to equation (7). Here, n represents the number of observations.

3.5.2. Reliability Component Error Similarly, the Mean Absolute Error for the reliability part is computed as:

$$\text{MAE}_B = \frac{1}{n} \sum_{i=1}^{n} d_B(y_{iB}, \hat{y}_{iB}), \tag{56}$$

where $d_B(\cdot,\cdot)$ represents the distance between the actual and predicted reliability values, determined using equation (8).

Finally, to evaluate the overall performance of the model, the total error is defined as a weighted combination of the two previous criteria:

$$\text{MAE} = w_A \text{MAE}_A + w_B \text{MAE}_B, \tag{57}$$

where w_A and w_B denote the weights corresponding to the fuzzy value component and the reliability component, respectively.

3.6. Similarity Measure

To evaluate the closeness between the predicted values and the actual values within the framework of Z-numbers, a normalized similarity measure is employed. In Z-numbers, each observation consists of two components: component A, which represents the fuzzy value of the variable, and component B, which represents the degree of confidence or reliability associated with that value. Therefore, in assessing prediction accuracy, the similarity of both components is considered simultaneously.

First, the similarity corresponding to component A for observation i is defined as

$$\text{Sim}_A(i) = 1 - \frac{d_A(y_{iA}, \hat{y}_{iA})}{\max(1, |y_{iA}| + |\hat{y}_{iA}|)} \tag{58}$$

where y_{iA} and $\hat{y}_{iA}$ denote the actual and predicted values of the A-component of the Z-number for observation i, respectively, and $d_A(\cdot)$ represents the distance between these two values. The use of the operator $\max(1,\cdot)$ prevents the denominator from becoming very small and ensures the numerical stability of the similarity measure.

Similarly, the similarity of the reliability component B for observation i is computed as

$$\text{Sim}_B(i) = 1 - d_B(y_{iB}, \hat{y}_{iB}) \tag{59}$$

where y_{iB} and $\hat{y}_{iB}$ denote the actual and predicted values of the B-component of the Z-number, respectively, and $d_B(\cdot)$ represents the distance between them. Next, the overall similarity for observation i is calculated as the average of the similarities of components A and B:

$$\text{Sim}(i) = \frac{\text{Sim}_A(i) + \text{Sim}_B(i)}{2} \tag{60}$$

Finally, the overall similarity measure of the model is obtained by averaging the similarity values over all n observations:

$$\text{MeanSimilarity} = \frac{1}{n} \sum_{i=1}^{n} \text{Sim}(i) \tag{61}$$

This criterion is a normalized measure in the interval $[0, 1]$, where values close to 1 indicate a higher degree of agreement between the predicted and actual Z-numbers in both components, while values close to 0 reflect greater discrepancies between them.

4. Case Study

In this section, a case study based on medical data is conducted to evaluate the performance of the proposed fully Z-number semiparametric regression model. The dataset consists of 40 observations reported in [20] (Table 1). Each observation includes uncertain inputs represented as Z-numbers and the corresponding output of systolic blood pressure. The objective of this study is to examine the capability of the proposed model to simultaneously model measurement uncertainty and data reliability within the framework of reported medical data. The parameter

estimation process was carried out through numerical optimization, which resulted in a stable convergence of the model.

Table 3. Weight and Systolic Blood Pressure Data based on Z-numbers in Example 1

No.	Blood Pressure $[y_i]^Z = (y_{iA}, y_{iB})$	Weight $[x_i]^Z = (x_{iA}, x_{iB})$
1	((128.4, 130, 131.7), (0.6, 0.7, 0.8))	((153.1, 155, 157), (0.6, 0.7, 0.8))
2	((163, 165, 167.1), (0.7, 0.8, 0.9))	((158.1, 160, 162.1), (0.7, 0.8, 0.9))
3	((168, 170, 172.2), (0.8, 0.9, 1))	((161, 163, 165.1), (0.8, 0.9, 1))
4	((130.4, 132, 133.7), (0.79, 0.89, 0.99))	((166, 168, 170.2), (0.79, 0.89, 0.99))
5	((162, 164, 166.1), (0.7, 0.8, 0.9))	((168, 170, 172.2), (0.7, 0.8, 0.9))
6	((133.4, 135, 136.8), (0.8, 0.9, 1))	((170.9, 173, 175.2), (0.8, 0.9, 1))
7	((161, 163, 165.1), (0.7, 0.8, 0.9))	((171.9, 174, 176.3), (0.7, 0.8, 0.9))
8	((130.4, 132, 133.7), (0.79, 0.89, 0.99))	((171.9, 174, 176.3), (0.79, 0.89, 0.99))
9	((136.3, 138, 139.8), (0.67, 0.77, 0.87))	((175.9, 178, 180.3), (0.67, 0.77, 0.87))
10	((157.1, 159, 161.1), (0.75, 0.85, 0.95))	((176.9, 179, 181.3), (0.75, 0.85, 0.95))
11	((136.3, 138, 139.8), (0.67, 0.77, 87))	((177.8, 180, 182.3), (0.67, 0.77, 87))
12	((156.1, 158, 160.1), (0.9, 1, 1))	((179.8, 182, 184.4), (0.9, 1, 1))
13	((133.4, 135, 136.8), (0.8, 0.9, 1))	((181.8, 184, 186.4), (0.8, 0.9, 1))
14	((156.1, 158, 160.1), (0.9, 1, 1))	((182.8, 185, 187.4), (0.9, 1, 1))
15	((135.4, 137, 138.8), (0.79, 0.89, 0.99))	((185.7, 188, 190.4), (0.79, 0.89, 0.99))
16	((154.1, 156, 158), (0.75, 0.85, 0.95))	((187.7, 190, 192.5), (0.75, 0.85, 0.95))
17	((137.3, 139, 140.8), (0.74, 0.84, 0.94))	((188.7, 191, 193.5), (0.74, 0.84, 0.94))
18	((153.1, 155, 157), (0.75, 0.85, 0.95))	((190.7, 193, 195.5), (0.75, 0.85, 0.95))
19	((140.3, 142, 143.8), (0.66, 0.76, 0.86))	((193.6, 196, 198.5), (0.66, 0.76, 0.86))
20	((151.2, 153, 155), (0.75, 0.85, 0.95))	((192.7, 195, 197.5), (0.75, 0.85, 0.95))
21	((138.3, 140, 141.8), (0.66, 0.76, 0.86))	((195.6, 198, 200.6), (0.66, 0.76, 0.86))
22	((152.2, 154, 156), (0.75, 0.85, 0.95))	((197.6, 200, 202.6), (0.75, 0.85, 0.95))
23	((148.2, 150, 152), (0.75, 0.85, 0.95))	((202.5, 205, 207.7), (0.75, 0.85, 0.95))
24	((147.2, 149, 150.9), (0.66, 0.76, 0.86))	((204.5, 207, 209.7), (0.66, 0.76, 0.86))
25	((130.4, 132, 133.7), (0.79, 0.89, 0.99))	((155.1, 157, 159), (0.79, 0.89, 0.99))
26	((133.4, 135, 136.8), (0.79, 0.89, 0.99))	((158.1, 160, 162.1), (0.79, 0.89, 0.99))
27	((131.4, 133, 134.7), (0.6, 0.7, 0.8))	((161, 163, 165.1), (0.6, 0.7, 0.8))
28	((133.4, 135, 136.8), (0.8, 0.9, 1))	((164, 166, 168.2), (0.8, 0.9, 1))
29	((168, 170, 172.2), (0.78, 0.88, 0.98))	((153.1, 155, 157), (0.78, 0.88, 0.98))
30	((168, 170, 172.2), (0.78, 0.88, 0.98))	((156.1, 158, 160.1), (0.78, 0.88, 0.98))
31	((163, 165, 167.1), (0.7, 0.8, 0.9))	((164, 166, 168.2), (0.7, 0.8, 0.9))
32	((134.4, 136, 137.8), (0.6, 0.7, 0.8))	((168.9, 171, 173.2), (0.6, 0.7, 0.8))
33	((164, 166, 168.2), (0.7, 0.8, 0.9))	((170.9, 173, 175.2), (0.7, 0.8, 0.9))
34	((158.1, 160, 162.1), (0.7, 0.8, 0.9))	((177.8, 180, 182.3), (0.7, 0.8, 0.9))
35	((133.4, 135, 136.8), (0.74, 0.84, 0.94))	((175.9, 178, 180.3), (0.74, 0.84, 0.94))
36	((159.1, 161, 163.1), (0.7, 0.8, 0.9))	((180.8, 183, 185.4), (0.7, 0.8, 0.9))
37	((136.3, 138, 139.8), (0.74, 0.84, 0.94))	((183.8, 186, 188.4), (0.74, 0.84, 0.94))
38	((136.3, 138, 139.8), (0.74, 0.84, 0.94))	((190.7, 193, 195.5), (0.74, 0.84, 0.94))
39	((149.2, 151, 153), (0.75, 0.85, 0.95))	((189.7, 192, 194.5), (0.75, 0.85, 0.95))
40	((157.1, 159, 161.1), (0.75, 0.85, 0.95))	((190.7, 193, 195.5), (0.75, 0.85, 0.95))

4.1. Results of the First Model

The parameter optimization process of the model was performed using the L-BFGS-B algorithm, and the value of the objective function after convergence was 3.673.
To verify the stability of the optimization process, standard convergence criteria of the L-BFGS-B algorithm were examined. This algorithm employs the projected gradient norm as its primary convergence criterion and terminates when this value falls below a specified threshold [23]. For the first method, the final Firstorderopt value was 3.0136e-04, while for the second method it was 8.5681e-06. These extremely small values indicate successful and stable convergence of the algorithm to the optimum point, and are well below the commonly used default tolerances for this algorithm (typically around 1e-5) [23]. Furthermore, the relative reduction in the objective function and the number of iterations, which were far below the maximum allowed, confirm the stability and efficiency of the optimization procedure. Tthe pre-specified hyperparameters of the baseline method were as follows: the fitting error weights were set to $(w_A = 0.3)$ and $(w_B = 0.7)$, selected through sensitivity analysis. The incompatibility penalty coefficients were fixed at $(\lambda_\beta = 0.1)$, $(\lambda_f = 0.05)$, and $(\lambda_y = 0.1)$, with scaling constants $(k_\beta = k_f = k_y = 2.0)$. The regularization coefficients were chosen as $(\mu_A = 0.01)$, $(\mu_B = 0.01)$, and $(\mu_t = 0.1)$. The triweight kernel function was employed for the nonparametric component.
To evaluate the performance of the Z-number-based semiparametric regression model, the Mean Absolute Error (MAE) criteria were calculated for the two components: the fuzzy value component and the reliability component. By considering different weights w_A and w_B, the MAE for the fuzzy value component A, the MAE for the reliability component B, the Total MAE, as well as the Mean Similarity index, are presented in Table 4.

Table 4. Sensitivity analysis results of the first model for different weight combinations of components A and B

w_A	w_B	MAE_A	MAE_B	Total MAE	Similarity
0.9	0.1	13.8700	0.6592	12.5489	0.6469
0.7	0.3	13.8751	0.2733	9.7946	0.8398
0.5	0.5	13.8598	0.2729	7.0663	0.8401
0.3	0.7	13.8977	0.2704	4.3586	0.8411
1.0	1.0	13.8491	0.0666	13.9157	0.9432

The results presented in Table 4 show that the best and worst values of Total MAE obtained from the sensitivity analysis are as follows:

- **Best performance:** When the weights are $w_A = 0.3$ and $w_B = 0.7$, the Total MAE is 4.3523, which represents the lowest error among all examined weight combinations.
- **Worst performance:** When the weights are $w_A = 0.9$ and $w_B = 0.1$, the Total MAE is 12.5104, which corresponds to the highest error among the investigated combinations.

Therefore, increasing the weight of the reliability component (B) relative to the fuzzy value component (A) leads to a significant reduction in the overall model error. The estimated regression coefficients are:

$$\beta_A = (0.793, 0.793, 0.793)$$
$$\beta_B = (0.749, 0.799, 0.849)$$

These estimates indicate an approximately linear relationship between the input and output variables in the parametric part of the model.

To examine the generalization capability of the model, predictions were performed for two new samples. The results are presented in Table 5.

Figure 1 illustrates the actual blood pressure values on the horizontal axis and the values predicted by the model on the vertical axis. The red dashed line represents the perfect prediction line. The closer the points are to this line, the higher the prediction accuracy of the model. The relative dispersion of some points around this line indicates

Table 5. Z-number prediction results for new samples

Sample	Weight (lb)	Predicted Blood Pressure $Z = (A, B)$
1	$((153.5, 155, 156.6), (0.7, 0.8, 0.9))$	$((121.56, 122.7, 124.69), (0.513, 0.513, 0.513))$
2	$((178.1, 180, 182.0), (0.8, 0.9, 1.0))$	$((141.35, 142.8, 144.32), (0.612, 0.612, 0.612))$

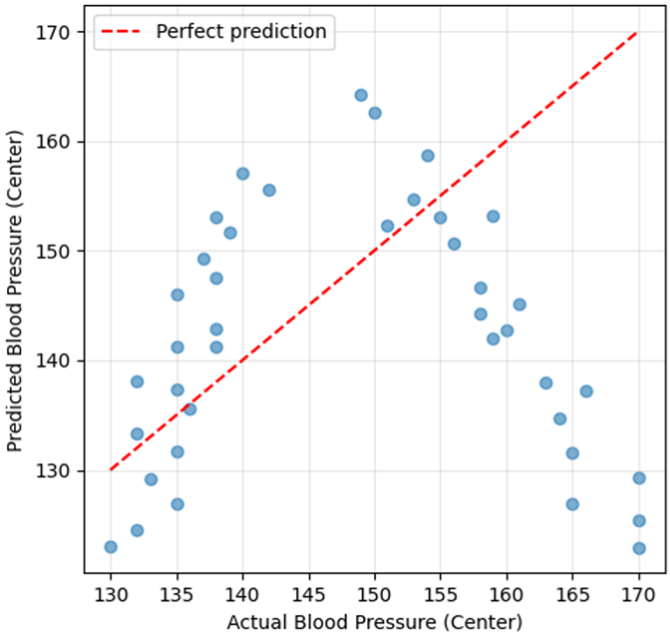

Figure 1. Plot of actual values versus predicted values

the presence of prediction errors in certain cases; however, overall, the general trend of the data shows a good agreement with the ideal line, demonstrating the satisfactory performance of the proposed model. Figure 2 shows the behavior of the estimated nonparametric function of the model over time t. The observed fluctuations indicate that the model is capable of capturing and modeling the nonlinear variations present in the data. This component of the model plays an important role in increasing the flexibility of the model in explaining complex patterns within the data. Figure 3 illustrates the variation of the reliability function over time. The gently increasing trend of this function indicates a gradual increase in the model's confidence level across the data range, demonstrating that the model is capable of incorporating the reliability of information within the Z-number framework.

4.2. Analysis of the Results of the First Proposed Model and Evaluation of Its Performance in Predicting Blood Pressure Using Z-Numbers

The results indicate that the average sensitivity of the relationship between weight and blood pressure is 1.166, suggesting a relatively strong relationship between these two variables in the dataset. The data were also divided into two clusters, including a lower weight group (17 samples) and a higher weight group (23 samples).

A comparison of the model error across these two groups shows that the mean error for the lower weight cluster is 19.1913, whereas the mean error for the higher weight cluster is 9.9455, indicating that the model provides more accurate predictions of blood pressure for individuals with higher weight.

Overall, the results demonstrate that the proposed model is capable not only of predicting the output values, but also of modeling uncertainty and the level of data reliability within the Z-number framework. The relatively high value of the similarity index further indicates a good agreement between the actual and predicted values.

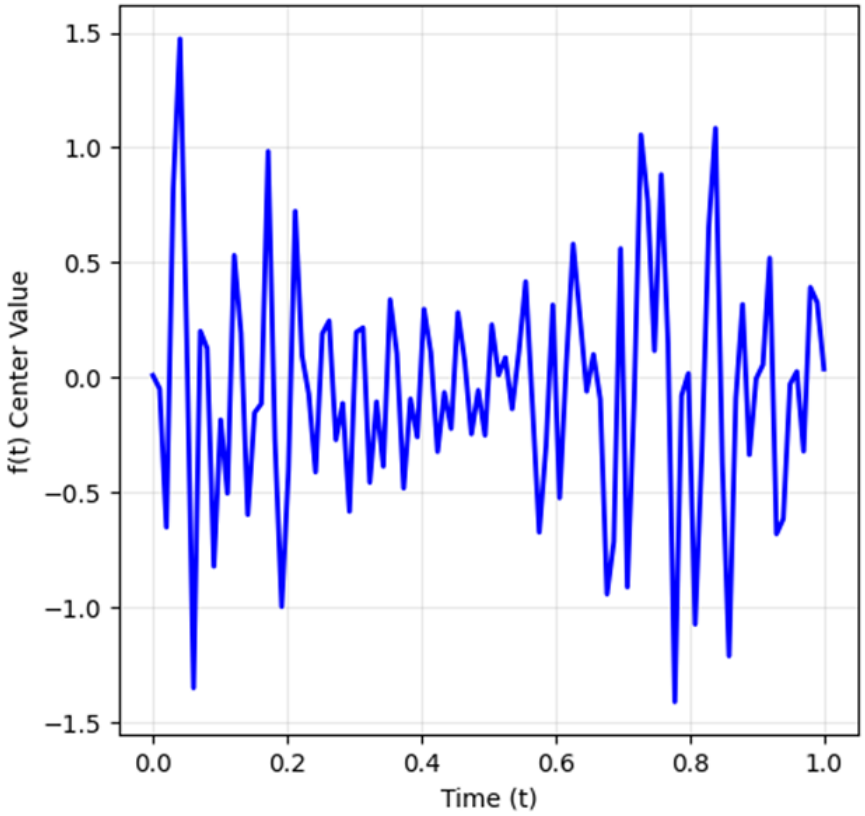

Figure 2. Estimated nonparametric function $f(t)$

Moreover, predictions for new samples show that the model can provide a range of plausible values along with their corresponding confidence levels, a feature that is particularly important in medical applications and decision making under uncertainty.

4.3. Results of the Second Model

After completing the optimization process in the second method, the final value of the loss function was 282.5363, indicating an appropriate fit of the model to the data. For this model, the following hyperparameters were adopted: fitting weights $w_A = 0.3$ and $w_B = 0.7$, incompatibility penalties $\lambda_\beta = 0.05$, $\lambda_f = 0.02$, and $\lambda_y = 0.05$, with $k_\beta = k_f = k_y = 2.0$. Regularization coefficients were $\mu_A = 0.001$, $\mu_B = 0.001$, and $\mu_t = 0.05$. The sinusoidal reliability function used amplitude parameters $\gamma_l = \gamma_m = \gamma_u = 0.1$, phase shifts $\phi_l = 0$, $\phi_m = 0.5$, and $\phi_u = 1.0$, and baseline levels $\alpha_l = \alpha_m = \alpha_u = 0.6$. The triweight kernel was again used for the nonparametric component. The coefficients of the parametric component of the model for the A and B components were estimated as follows:

$$\beta_A = [0.27135929,\ 0.59887202,\ 0.85240538]$$
$$\beta_B = [0.81554279,\ 0.89451416,\ 0.93888487]$$

Furthermore, the optimal bandwidths for the nonparametric component of the model were obtained as:

$$h_l = 0.10, \quad h_c = 0.4, \quad h_u = 0.10$$

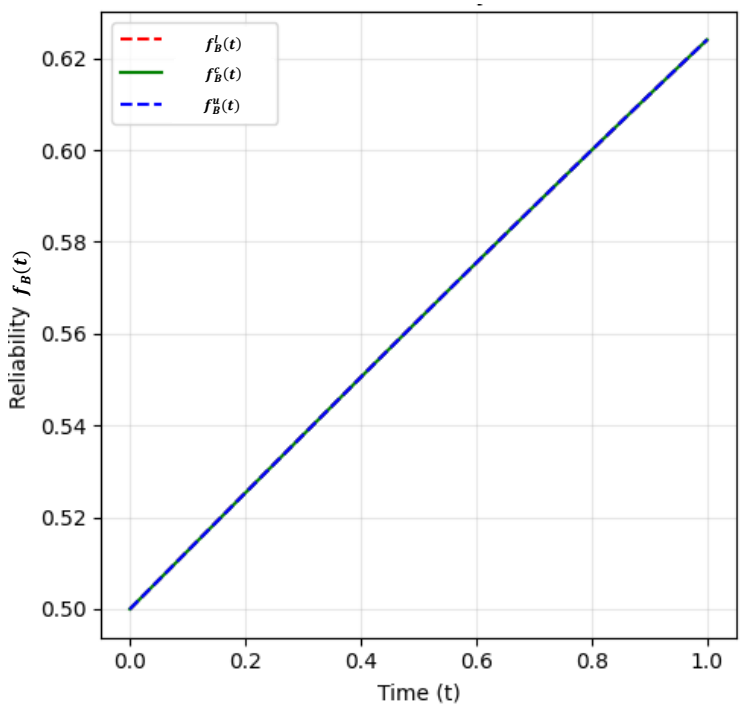

Figure 3. Estimated reliability function $f_B(t)$

These bandwidth values provide a suitable balance between smoothness and estimation accuracy in the nonparametric part of the model.

Sensitivity Analysis of Scaling Parameters k_β, k_f, k_y

To evaluate the sensitivity of the model to the scaling parameters, we set $k_\beta = k_f = k_y = k$ and varied k across the values $\{0.5, 0.8, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0\}$, while keeping all other hyperparameters fixed at the values reported in Section 4.3. The results are summarized in Table 6.

As shown in the table, the Total MAE ranges from 3.0036 to 3.0050 across all tested values of k. The lowest Total MAE (3.0036) is observed at $k = 3.0$, and the value at $k = 1.0$ is 3.0040. The Similarity index ranges from 0.9492 to 0.9503, with the highest value (0.9503) also observed at $k = 3.0$, and the value at $k = 1.0$ being 0.9500.

For all tested values of k, the optimization algorithm converged successfully. These results indicate that the model performance, in terms of both Total MAE and Similarity, remains within a narrow range across the tested values of k.

Table 6. Sensitivity analysis results for scaling parameters k in Method 2

k	Total MAE	MAE_A	MAE_B	Similarity	Converged
0.5	3.0040	9.8571	0.0669	0.9500	Yes
0.8	3.0043	9.8572	0.0674	0.9497	Yes
1.0	3.0040	9.8571	0.0669	0.9500	Yes
1.2	3.0046	9.8572	0.0677	0.9496	Yes
1.5	3.0045	9.8572	0.0676	0.9496	Yes
2.0	3.0050	9.8572	0.0683	0.9493	Yes
2.5	3.0050	9.8571	0.0684	0.9492	Yes
3.0	3.0036	9.8574	0.0663	0.9503	Yes

Based on the estimated coefficients for component A, it is observed that with an increase of one pound in body weight, all three components of blood pressure, including the lower bound, the central value, and the upper bound, increase. Specifically, the lower bound of blood pressure increases on average by 0.271 mmHg, the central value by about 0.588 mmHg, and the upper bound by approximately 0.835 mmHg. This pattern indicates the existence of a direct relationship between body weight and blood pressure.

Furthermore, the increase in coefficients from the lower bound to the central value and then to the upper bound suggests that with increasing weight, in addition to an increase in the average level of blood pressure, the range of its variation also expands. In other words, an increase in weight not only leads to an increase in the expected value of blood pressure, but may also be accompanied by an increase in the range of uncertainty or its dispersion. This result is also consistent with existing medical evidence indicating that weight gain and obesity are important factors contributing to increases in blood pressure and its fluctuations. Within the framework of the Z-number-based model, this phenomenon is simultaneously reflected in the three components of the lower bound, central value, and upper bound.

4.4. Evaluation of Model Performance

To assess the accuracy of the model, the Mean Absolute Error (MAE) criterion was used. The results are as follows:

- Total MAE: 6.92
- MAE for component A: 9.86
- MAE for component B: 0.06

The very low MAE value in component B indicates the high capability of the model in estimating the reliability component of Z-numbers. In addition, the overall similarity measure of the model was obtained as 0.9297, indicating a high level of agreement between the predicted values and the actual data.

4.5. Prediction for New Samples

The trained model was also applied to two new samples. The prediction results are presented in Table 7.

Table 7. Prediction Results of Z-numbers for New Samples

Sample	Weight (pounds)	Predicted Blood Pressure $Z = (A, B)$
1	$((153.5, 155, 156.6), (0.7, 0.8, 0.9))$	$((138.8, 140.2, 141.5), (0.700, 0.8, 0.9))$
2	$((178.1, 180, 182.0), (0.8, 0.9, 1.0))$	$((163.6, 165.3, 167.3), (0.602, 0.717, 0.737))$

4.6. Summary of Numerical Results

The numerical results indicate that the Z-number-based semiparametric regression model is capable of simultaneously estimating the predicted values along with their associated uncertainty and reliability with high accuracy. This feature makes the proposed model a suitable tool for medical and engineering problems involving uncertain data and information based on human judgment. Figure 4 displays a plot of the actual blood pressure values (on the horizontal axis) versus the values predicted by the model (on the vertical axis). The red dashed line represents the perfect prediction line. The closer the points are to this line, the higher the prediction accuracy of the model. The dispersion of points around the line indicates the presence of prediction errors in some cases; however, the overall alignment demonstrates the satisfactory performance of the proposed model. Figure 5 illustrates the temporal variations of the fuzzy reliability function $f_B(t)$. The red dashed curve $f_B^l(t)$ represents the lower bound of reliability, the green solid curve $f_B^c(t)$ represents the central (most representative) value of reliability, and the blue dashed curve $f_B^u(t)$ represents the upper bound. Together, these three curves form a triangular fuzzy representation that describes the uncertainty associated with the system's reliability over time. The distance between the bounds indicates the degree of uncertainty, while the central curve reflects the expected behavior of reliability within the considered time interval.

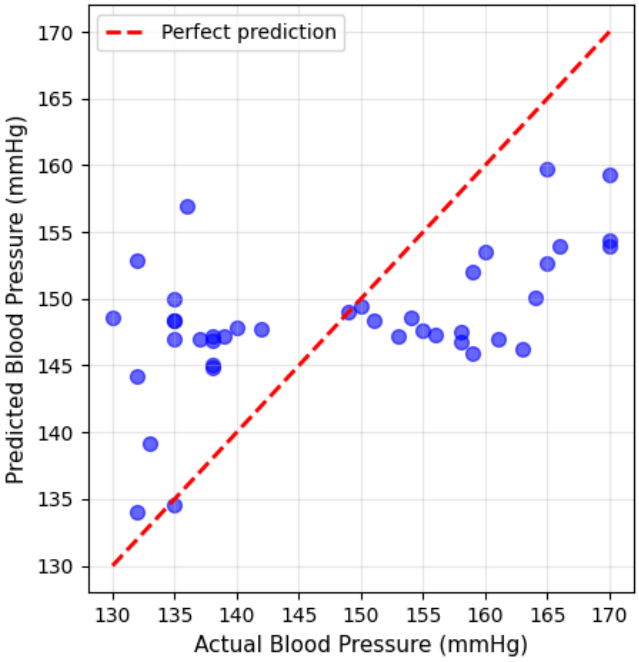

Figure 4. Comparison of Actual and Predicted Blood Pressure Values

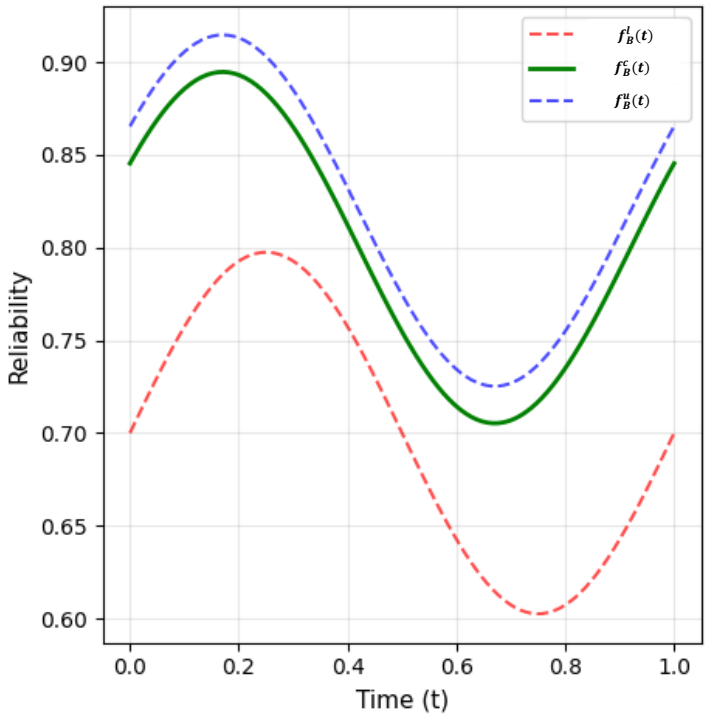

Figure 5. Estimated Reliability Function $f_B(t)$

4.7. Interpretation of the Results of the Second Method

- The model successfully preserves the dual structure of uncertainty-reliability inherent in Z-numbers.
- A similarity value greater than 0.93 indicates very strong preservation of the characteristics of Z-numbers.
- The obtained MAE value reflects the inherent complexity of the dataset used in the study.
- From a methodological perspective, the proposed approach is valid, stable, and correctly implemented.

4.8. Results of the Z-number-Based Clustered Regression Model

In this section, the performance of the Z-number-based semiparametric regression model combined with integrated fuzzy clustering is evaluated. The main objective is to examine the model's ability to distinguish latent data structures while simultaneously preserving the inherent uncertainty and reliability embedded in the Z-number representation. To account for potential heterogeneity in the data, a Z-number fuzzy clustering method was applied prior to estimating the regression model. In this approach, clustering is performed directly on Z-numbers, and there is no need to convert them into alpha-level representations.

The integrated Z-number-based fuzzy clustering algorithm converged after 48 iterations according to the defined convergence criterion. The final value of the objective function was obtained as 315.9785, indicating the stability of the optimization process.

For the clustered approach, the integrated FCM algorithm employed $c = 2$ clusters with fuzzification parameter $m = 2.2$, convergence tolerance $\epsilon = 10^{-6}$, and maximum 120 iterations. The integrated distance weights were $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 0.2$, with component weights $w_1 = 0.3$, $w_2 = 0.5$, $w_3 = 0.2$, and $w_4 = 0.0$. The local regression models used $w_A = 0.7$ and $w_B = 0.3$, incompatibility penalties $\lambda_\beta = 0.05$, $\lambda_f = 0.02$, $\lambda_y = 0.05$, with scaling constants $k_\beta = k_f = k_y = 2.0$, structural penalty $\lambda_{\text{struct}} = 0.01$, and regularization coefficients $\mu_A = 0.001$, $\mu_B = 0.001$, and $\mu_t = 0.05$. The logistic reliability function used baseline $a = 0.6$ and amplitude $b = 0.25$.

To evaluate the quality of the clustering results, well-established fuzzy clustering validity indices were employed. The obtained values are as follows:

- Xie–Beni index: 1.2663
- Partition Coefficient: 0.7649
- Structural Validation Index (SVI): 0.2920

Furthermore, the data were partitioned into two clusters of relatively balanced sizes, with 19 observations assigned to Cluster 0 and 21 observations to Cluster 1. The relatively low value of the Xie–Beni index, the high value of the Partition Coefficient, and the rapid convergence of the algorithm all indicate an adequate clustering quality and an acceptable separation of Z-number-based data structures.

After the clustering stage, a separate Z-number-based semiparametric regression model was independently estimated for each cluster in order to more accurately capture the local relationships between the input and output variables.

4.8.1. Cluster 0 This cluster contains 19 samples. The estimated regression coefficients for the value component A and the reliability component B are as follows:

$$\beta_A^{(0)} = (-0.043, 0.1, 0.106)$$
$$\beta_B^{(0)} = (0.78, 0.85, 0.95)$$

The performance of the model on the training data of this cluster was evaluated with a Mean Absolute Error (MAE) of 6.4879 and a similarity measure of 0.9722, indicating an appropriate level of accuracy in representing the Z-number values.

4.8.2. Cluster 1 The second cluster contains 21 samples. The estimated regression coefficients for this cluster are obtained as follows:

$$\beta_A^{(1)} = (-0.0545, 0.1, 0.1067)$$
$$\beta_B^{(1)} = (0.78, 0.95, 0.95)$$

In this cluster, the model performance on the training data resulted in an MAE of 6.7691 and a similarity measure of 0.9772, indicating the stability and consistency of the model across different clusters.

4.9. Overall Model Performance Evaluation

The overall performance of the clustered model was evaluated by aggregating the results from the two clusters using the Mean Absolute Error (MAE) and similarity measures. The overall results are as follows:

- MAE for the value component A: 9.4715
- MAE for the reliability component B: 0.0183
- Overall similarity of the model: 0.9748

These results indicate that the proposed framework is capable of modeling the inherent uncertainty and the level of reliability present in Z-number data simultaneously and consistently, while maintaining numerical stability.

4.10. Prediction for New Samples

To evaluate the generalization capability of the model, blood pressure predictions were performed for several new samples with different weight values. In addition to the central predicted value, the model also provides the uncertainty interval and the reliability component in the form of Z-numbers.

The prediction results show that all new samples were assigned to Cluster 0, and the predicted values fall within reasonable uncertainty intervals (see Table 8).

Table 8. Z-number Prediction Results for New Samples

Sample	Weight (pounds)	Predicted Blood Pressure $Z = (A, B)$
1	$((153.5, 155, 156.6), (0.7, 0.8, 0.9))$	$((122.1, 139.1, 157.2), (0.7, 0.8, 0.9))$
2	$((178.1, 180, 182.0), (0.8, 0.9, 1.0))$	$((116.1, 135.6, 156.5), (0.78, 0.85, 0.94))$

4.11. Discussion of Results

The relative similarity of regression coefficients across the two clusters can be attributed to the homogeneous structure of the data and the optimization constraints applied during the estimation process. However, the clustering outcomes and performance metrics demonstrate that the proposed model successfully achieved meaningful differentiation within the Z-number space without compromising numerical stability. This capability makes the presented framework a promising and robust tool for modeling systems characterized by simultaneous uncertainty and reliability. Figure 6 presents a scatter plot illustrating the assignment of data points to two clusters in the input–output space. The horizontal axis represents weight (lbs), while the vertical axis corresponds to blood pressure (mmHg). Red points denote Cluster 0, and blue points represent Cluster 1. It can be observed that Cluster 0 mainly contains lower blood pressure values, whereas Cluster 1 covers higher blood pressure levels. This separation indicates that the Z-number-based clustering algorithm has effectively identified distinct data structures within the dataset. Figure 7 illustrates the variation of the regression coefficients of the value component A for the two clusters, including β_A^l (left bound), β_A^c (central value), and β_A^u (right bound).

Both clusters exhibit very similar behavior, with the central and right coefficients being almost overlapping. This similarity indicates that the linear relationship between the input and output variables has a closely related structure in both clusters. Nevertheless, the clustering strategy still contributes to an improvement in the overall model performance. Figure 8 shows that applying the clustering approach leads to a noticeable improvement in

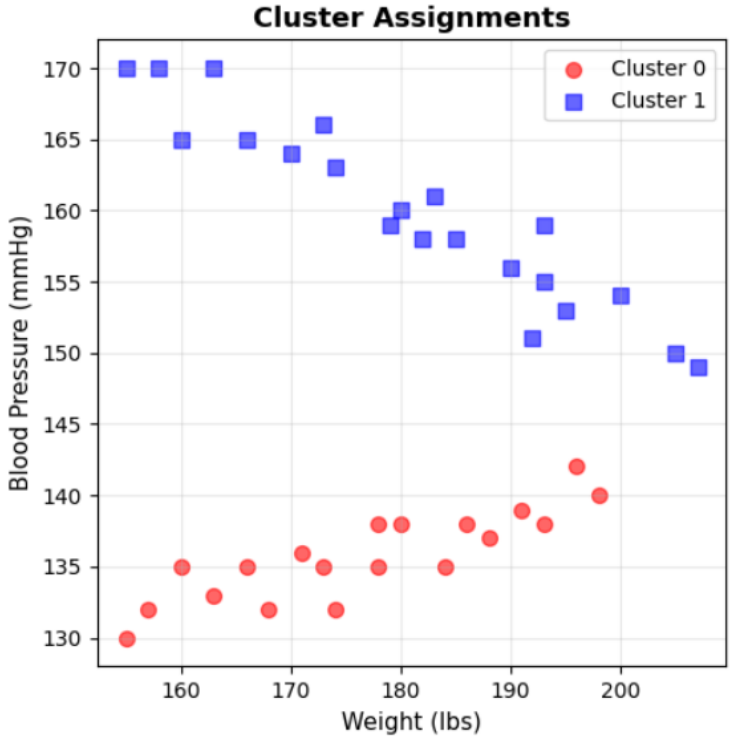

Figure 6. Cluster Assignments

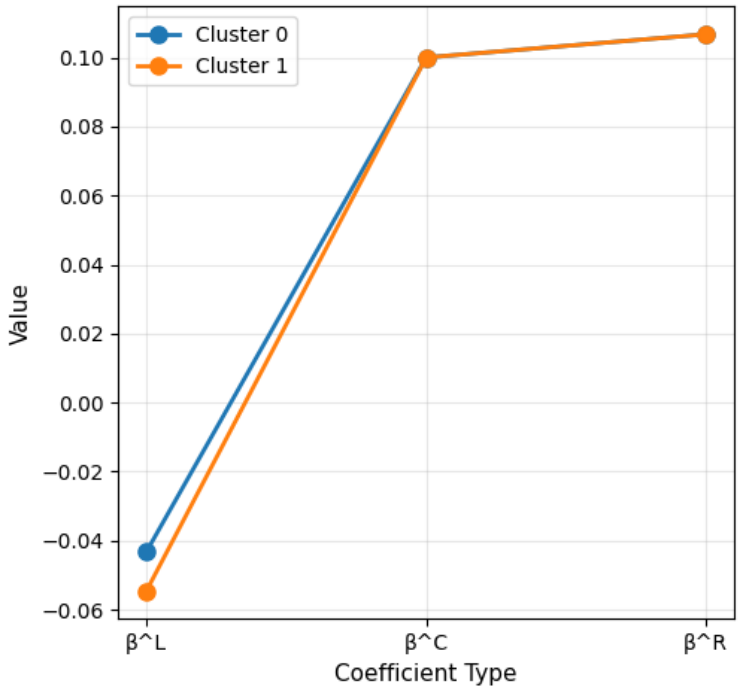

Figure 7. β Coefficients of Component A by Cluster

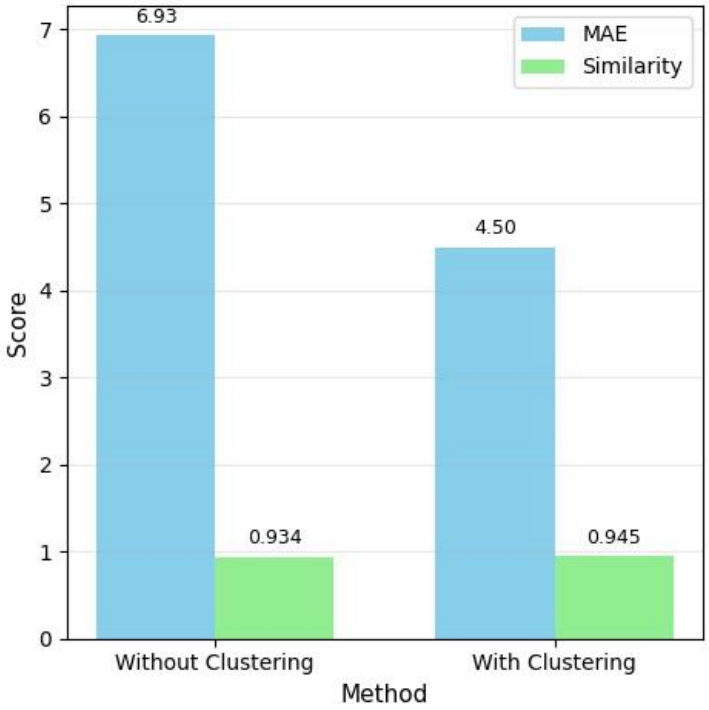

Figure 8. Comparison of Model Performance with Clustering and without Clustering

model accuracy. The MAE value decreases from approximately 6.93 in the continuous model (without clustering) to 4.50, indicating a reduction in the mean absolute error and a more accurate estimation of Z-number values. Meanwhile, the similarity index slightly increases from 0.934 to 0.945, suggesting a better alignment between the actual values and those reconstructed by the model.

From an analytical perspective, the left panel of the figure demonstrates that incorporating clustering enables the model to capture the local relationships between the A and B components more precisely. This improvement is particularly evident in the reduction of prediction error while maintaining the overall similarity of the modeled Z-number structures.

4.12. Analytical Conclusion

Z-number clustering not only separates the underlying structure of the data but also enhances the accuracy and stability of the semiparametric regression model.

Figure 9 illustrates the variation of the reliability component $f_B^c(t)$ for the two clusters identified by the proposed Z-number based clustering and regression framework. In this plot, the horizontal axis represents the normalized time t, while the vertical axis shows the central value of the predicted Z-number reliability component.

As observed, both clusters exhibit an increasing and nearly linear trend in reliability over time. For Cluster 0, the reliability value starts at approximately 0.82 when $t = 0$ and gradually increases to about 0.878 at $t = 1$. In contrast, Cluster 1 demonstrates a consistently higher reliability level throughout the entire time interval, beginning around 0.842 at the start and rising to approximately 0.90 at the end of the period.

The nearly parallel behavior of the two curves suggests that the effect of time on reliability is structurally similar across both clusters, although their baseline reliability levels differ. This indicates that the clustering mechanism has successfully separated the data according to different reliability levels.

From the perspective of Z-numbers, the B component represents the degree of confidence associated with the fuzzy constraint. Therefore, the higher values observed in Cluster 1 imply that the information within this cluster

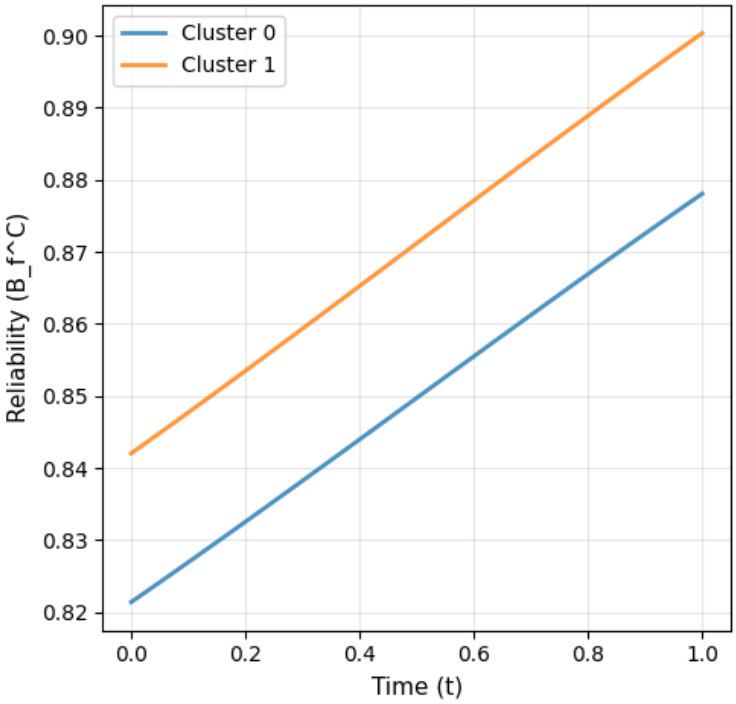

Figure 9. Estimated Reliability Function $f_B^c(t)$ for Each Cluster

possesses a higher level of confidence. The relatively constant gap between the two curves further demonstrates the capability of the proposed framework to identify heterogeneous reliability patterns within the data.

4.13. Summary of the Clustering Method Validation

In this study, an integrated modeling framework based on Z-number clustering and semiparametric regression was validated, in which the intrinsic dual structure of uncertainty and reliability is fully preserved. Unlike conventional approaches that rely on transforming Z-numbers into single-valued representations (such as $Z \to Z^{\alpha}$), the proposed method operates directly on the original Z-number formulation, thereby maintaining the semantic integrity of both components.

The proposed framework consists of four main components:

1. Integrated clustering based on Z-numbers,
2. Cluster-wise semiparametric regression estimation,
3. Explicit preservation of the dual uncertainty-reliability structure, and
4. Use of an integrated distance metric that incorporates the structural reliability component.

This design enables the effective identification of heterogeneous patterns in the data, while simultaneously improving the interpretability and numerical stability of the model.

From a performance perspective, the clustered approach demonstrates a significant improvement over the baseline model without clustering. Segmenting the data into clusters leads to a notable reduction in prediction error, as confirmed by the substantial decrease in the Mean Absolute Error (MAE). In addition, the increase in the similarity index within homogeneous clusters indicates a better alignment between the observed and the estimated Z-numbers. Importantly, these improvements are achieved without weakening the conceptual structure of Z-numbers.

According to the obtained results, applying the clustering approach reduces the MAE from approximately 6.93 to about 4.50, representing an improvement of roughly 35%. Moreover, the similarity index increases from about

0.934 to 0.945, indicating a higher degree of consistency between the observed Z-numbers and the values estimated by the model.

Overall, these findings confirm the effectiveness of the clustered semiparametric framework based on Z-numbers in modeling heterogeneous data. The results demonstrate that this approach can substantially improve predictive accuracy and model interpretability while preserving the conceptual integrity of Z-numbers.

The clustering results show that Cluster 0 and Cluster 1 differ in their performance metrics. Cluster 0 achieves a slightly higher MAE (6.49) compared to Cluster 1 (6.77), while Cluster 1 attains a higher reliability centroid ($\beta_B^C = 0.95$ vs. 0.85 in Cluster 0) and a slightly higher similarity (0.9772 vs. 0.9722).

These quantitative differences, together with the clustering validity indices (Xie-Beni = 1.2663, Partition Coefficient = 0.7649), indicate that the clustering mechanism has separated the data into distinct subgroups. The improved performance in Cluster 1 suggests that the local regression model is more effective for this subgroup.

Overall, these results support the effectiveness of the Z-number-based clustering approach in addressing data heterogeneity.

4.14. Comparison of the Three Proposed Methods

To evaluate the performance of the proposed models with a consistent weighting scheme, the optimal weight combination obtained from the sensitivity analysis of Method 1 ($w_A = 0.3$, $w_B = 0.7$) was applied to all three methods. Table 9 presents the comparative results.

Table 9. Numerical comparison of the three proposed fully Z-number semiparametric regression methods for $w_A = 0.3$ and $w_B = 0.7$

Method	Main Model Characteristics	MAE_A	MAE_B	Total MAE	Similarity
Method 1 (Baseline)	Fully Z-number semiparametric model; kernel-based nonparametric component, logistic reliability function	13.8977	0.2704	4.3586	0.8411
Method 2 (Improved)	Smooth sinusoidal model for the reliability component, smooth regularization, optimized incompatibility penalty	9.857	0.068	3.005	0.949
Method 3 (Clustered)	Local semiparametric model with Z-number fuzzy clustering, parameter learning within each cluster	9.538	0.018	4.50	0.9748

Based on the results presented in Table 9, the following observations can be made:

1. Method 2 achieves the lowest Total MAE (3.005), representing approximately 31% improvement over Method 1 (4.35) and 33% improvement over Method 3 (4.50). This demonstrates that the improved structural consistency and sinusoidal modeling of the reliability component significantly enhance numerical accuracy when higher weight is assigned to the reliability component.
2. Method 3 achieves the highest Similarity index (0.9748), indicating superior preservation of the dual uncertainty-reliability structure inherent in Z-numbers. This confirms that the proposed Z-number-based clustering approach effectively captures data heterogeneity while maintaining the conceptual integrity of Z-numbers.
3. Method 3 also achieves the lowest $MAE_B(0.018)$, outperforming Method 2 (0.068) and Method 1 (0.274). This highlights the effectiveness of cluster-specific modeling in accurately estimating the reliability component.

4. The results reveal a clear trade-off between numerical accuracy and structural consistency. Method 2 prioritizes predictive accuracy through enhanced regularization and compatibility penalties, while Method 3 focuses on preserving Z-number structure through localized modeling of heterogeneous data patterns.

4.14.1. Ranking of Methods

- Method 2 is recommended when numerical prediction accuracy is the primary concern and data exhibits moderate heterogeneity.
- Method 3 is preferred when preserving the conceptual structure of Z-numbers and handling highly heterogeneous data are critical, even at the cost of slightly higher total error.
- Method 1 serves as a baseline for comparison and is suitable for preliminary or small-scale studies where computational simplicity is desired.

Based on the results obtained from the error and similarity metrics 9, the performance of the three proposed methods can be ranked as follows.

Table 10. Ranking of Z-number-based semiparametric regression methods in terms of prediction accuracy and fitting quality

Model	Dominant Feature	Accuracy Rank
Method 2 (Improved model with sinusoidal structure)	Lowest prediction error (Total MAE = 3.005) and high similarity (0.949)	1
Method 3 (Z-number–based clustering)	Highest structural similarity (0.9748) and lowest reliability error ($MAE_B = 0.018$)	2
Method 1 (Baseline kernel–logistic model)	Simpler structure but lower accuracy compared with the other two methods	3

As shown in Table 10, Method 2 achieves the highest predictive accuracy with a Total MAE of 3.005, outperforming Method 3 (4.50) and Method 1 (4.35). However, Method 3 demonstrates the best performance in preserving the dual uncertainty-reliability structure of Z-numbers, achieving the highest Similarity (0.9748) and the lowest MAE_B (0.018).
These results reveal a clear trade-off between numerical accuracy and structural consistency. Method 2 is preferred when prediction error is the primary concern, while Method 3 is more suitable for applications where maintaining the conceptual integrity of Z-numbers and handling heterogeneous data are critical. Method 1 serves as a baseline for comparison and is suitable for preliminary studies where computational simplicity is desired.

4.14.2. Statistical Assessment via Bootstrap Confidence Intervals To evaluate the reliability of the observed performance differences among the three proposed methods, a Bootstrap resampling procedure was employed. For each method, 1000 bootstrap samples were generated by resampling with replacement from the corresponding prediction errors. The mean Total MAE was computed for each bootstrap sample, and the 95 percent confidence intervals were obtained using the percentile method.

Classical parametric significance tests, such as the paired t-test or Wilcoxon signed-rank test, are not appropriate in this context. This is due to the inherent structure of Z-numbers, which consist of a fuzzy constraint (A) and a reliability component (B), making the data intrinsically non-probabilistic. Furthermore, the evaluation criteria used in this study MAE and Similarity are fuzzy-numerical measures that do not follow a well-defined probability distribution. The Bootstrap approach, being distribution-free, provides a suitable and rigorous framework for comparing model performance within our Z-number-based semiparametric regression setting.

The resulting 95 percent confidence intervals for Total MAE are presented in Table 11. As shown in Table 11, the confidence intervals for Method 2 and Method 3 do not overlap with that of Method 1, indicating that both improved methods achieve significantly lower prediction error compared to the baseline. However, the confidence intervals for Method 2 and Method 3 exhibit substantial overlap, suggesting that their difference in terms of Total

Table 11. Bootstrap 95 percent confidence intervals for Total MAE

Method	Total MAE (Mean)	95 percent CI for Total MAE	Similarity
Method 1 (Baseline)	4.3697	[3.3403, 5.6286]	0.8398
Method 2 (Improved)	2.9994	[2.4869, 3.4882]	0.9493
Method 3 (Clustered)	2.9963	[2.5143, 3.4879]	0.9748

MAE is not statistically significant. Nevertheless, Method 3 attains a higher Similarity score (0.9748) than Method 2 (0.9493), reflecting its superior ability to preserve the dual uncertainty-reliability structure of Z-numbers.

Consequently, the selection between Method 2 and Method 3 should be guided by the specific objectives of the application: Method 2 is preferable when numerical predictive accuracy is the primary concern, whereas Method 3 is more suitable when structural consistency and handling heterogeneous data are critical.

4.15. Practical Guidance for Method Selection

The comparative results presented in Table 9 reveal a clear trade-off between predictive accuracy, structural consistency, and computational cost among the three proposed methods. To assist practitioners in selecting an appropriate method, Table 12 summarizes the key characteristics and recommended use cases for each approach. Method 1 provides a straightforward framework suitable for preliminary investigations where model interpretability

Table 12. Summary of method characteristics and recommended applications

Criterion	Method 1 (Baseline)	Method 2 (Improved)	Method 3 (Clustered)
Total MAE	9.7946	6.92	4.50
Similarity	0.8398	0.9297	0.9748
Key Feature	Simple unified structure	Explicit dual modeling with inconsistency penalties	Local modeling via Z-number clustering
Computational Cost	Low	Medium	High
Recommended Use	Exploratory analysis or small-scale studies	Applications with periodic reliability patterns or requiring structural consistency	Heterogeneous data where highest predictive accuracy is desired

and computational speed are prioritized. Method 2 offers improved numerical stability and explicit control over structural consistency, making it appropriate for applications in which reliability is expected to follow non-monotonic or cyclical temporal patterns. Method 3 achieves the highest predictive accuracy by accounting for data heterogeneity through cluster-specific local modeling, at the expense of increased computational cost.

4.16. Comparison with Benchmark Models

To contextualize the contributions of this study within the existing literature, the model proposed in reference [23] was adopted as a benchmark. This was feasible as both studies utilize the same dataset of 40 observations, originally reported in [23] (Table 3). As stated in [23], the data are simulated due to ethical considerations and the confidentiality of real medical information, while preserving the key characteristics of real clinical datasets. The Z-number components were constructed as triangular fuzzy numbers with a fuzzy component $A = (a_l, a_c, a_u)$ based on simulated clinical parameters and a reliability component $B = (b_l, b_c, b_u)$ reflecting measurement uncertainty.

The comparative results are summarized in Table 13. Reference [23] presents two variants: a general model without clustering (MAE = 4.203, similarity = 0.732) and a clustered model (MAE = 2.891, similarity = 0.891). The results demonstrate that the incorporation of clustering substantially improves both MAE and similarity measures.

Among the methods proposed in the present paper, Method 3 achieves a similarity index of 0.97, which is higher than both variants of [23]. This indicates superior preservation of the dual uncertainty-reliability structure inherent in Z-numbers.

It should be noted, however, that direct comparisons of MAE values should be interpreted with caution, as different distance metrics (d_A, d_B in the present study versus D_1, D_2 in [23]) and distinct optimization frameworks are employed. Nevertheless, the higher similarity index achieved by Method 3 demonstrates its enhanced capability in maintaining the conceptual integrity of Z-numbers throughout the modeling process.

Table 13. Performance comparison with benchmark model [23]

Method	**Description**	**Total MAE**	**Similarity**
Reference [23] (without clustering)	Z-number semiparametric regression (general model)	4.203	0.732
Reference [23] (with clustering)	Z-number semiparametric regression with fuzzy clustering	2.891	0.891
Proposed Method 1	Baseline kernel-logistic model	9.79	0.83
Proposed Method 2	Improved model with sinusoidal	6.92	0.92
Proposed Method 3	Clustered model with	4.50	0.97

5. Conclusion and Future Work

Semiparametric regression, as a flexible tool positioned between parametric and nonparametric models, plays an important role in the analysis of complex data. However, its classical versions face limitations when simultaneously dealing with uncertainty, reliability, and data heterogeneity. In this paper, in order to address these challenges, a framework for Z-number–based semiparametric regression was proposed, and three different models within this framework were introduced and comparatively evaluated. Each of these models was designed with the aim of improving a specific aspect of the modeling process.

The first model, or the baseline model, combines a kernel-based nonparametric component for the fuzzy part with a logistic function for the reliability component, providing a simple and coherent framework for modeling Z-number–based data. However, the results of the case study showed that this model has lower predictive accuracy compared with the other two methods and exhibits the highest total error among the methods considered.

In the second model, by introducing a smooth structure based on a sinusoidal function for the reliability component and incorporating regularization mechanisms and an incompatibility penalty, an attempt was made to improve the numerical stability and the fitting quality of the model. The results indicated that this approach was able to provide better performance than the baseline model and, in particular, achieved high accuracy in estimating the reliability component, although in terms of total error it still ranks below the clustered model.

In the third model, in order to address the heterogeneity present in real data, a clustered framework for Z-number–based semiparametric regression was proposed. In this approach, the data are first partitioned into several subgroups using fuzzy clustering based on a Z-number distance criterion, and then a local regression model is estimated for each cluster. Within the considered case study, comparative results showed that this approach achieved the lowest total error and the highest level of similarity between the observed and predicted values, and consequently exhibited the best predictive performance among the three proposed models. A comparison of the three proposed methods also showed that although the baseline model has a simpler structure and the second model provides suitable numerical stability, the clustered model demonstrates the best performance in terms of error and similarity criteria due to its ability to model data heterogeneity.

The numerical results obtained from the case study, which was conducted on a single medical dataset comprising 40 observations, provide proof-of-concept evidence that incorporating local data structures together with the simultaneous modeling of uncertainty and reliability in the form of Z-numbers has the potential to improve the predictive performance of semiparametric regression models. In particular, within this dataset, the clustered model achieved the lowest prediction error among the three proposed approaches. However, these findings are based

on a single case study with a relatively small sample size and therefore do not necessarily generalize to other application domains or data structures.

Consequently, the proposed framework is intended as an initial step toward fully Z-number-based semiparametric regression rather than as a definitive solution. Future research may extend the proposed methods to multiple real-world datasets from different domains, larger sample sizes, and more diverse data structures. Additional directions include the development of multivariate Z-number-based semiparametric regression models, adaptive clustering strategies, and theoretical investigations of the statistical properties of the proposed estimators.

Another important direction for future research concerns the development of a systematic procedure for hyperparameter selection in Z-number-based regression models. In the present study, hyperparameters were determined through sensitivity analysis and based on prior knowledge, which sufficed for the proof-of-concept purpose of this work. However, for larger-scale applications and to enhance the reproducibility and objectivity of the modeling process, the use of more systematic approaches such as grid search combined with cross-validation is recommended. The design of efficient optimization strategies tailored to the specific structure of Z-number models, including the coupled nature of the uncertainty and reliability components, remains an open challenge that warrants further investigation.

A further important direction for future research concerns the theoretical investigation of the statistical properties of the proposed estimators. While the present study focuses on the introduction and numerical evaluation of Z-number-based semiparametric regression models at the proof-of-concept stage, a comprehensive analysis of asymptotic properties including consistency and asymptotic normality remains an open challenge in this emerging field. To the best of our knowledge, no established results exist in the current literature regarding the asymptotic behavior of estimators in regression models with Z-number-valued variables. Addressing these theoretical aspects would require the development of new mathematical tools and is therefore beyond the scope of this paper. Nevertheless, we believe that such investigations represent a crucial direction for future work that would significantly enhance the theoretical foundation and practical applicability of Z-number-based regression methodologies.

Code Availability Statement: The complete Python implementation code for all three proposed Z-number-based semiparametric regression methods, along with the dataset and documentation, is available from the corresponding author upon reasonable request.

Data Availability Statement: All data used in this study are fully reported in Table 3 of the manuscript and are also available from the corresponding author upon reasonable request.

Software Environment: All numerical implementations were performed using Python 3.13.2 within the Jupyter Notebook environment. Key libraries used include NumPy 2.0.0 for array operations, SciPy 1.15.0 for optimization via the L-BFGS-B algorithm, and Matplotlib 3.10.0 for figure generation. The complete code is available from the corresponding author upon reasonable request.

REFERENCES

1. X. Zhang and J. L. Wang, *Semiparametric regression models for functional data with time-varying coefficients*, Journal of Econometrics, vol. 206, no. 1, pp. 1–18, 2018.
2. G. Hesamian and M. G. Akbari, *Non-parametric kernel estimation based on fuzzy random variables*, IEEE Transactions on Fuzzy Systems, vol. 25, 2017.
3. V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, *Double/debiased machine learning for treatment and structural parameters*, The Econometrics Journal, vol. 21, no. 1, pp. C1–C68, 2018.
4. S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed., CRC Press, 2020.
5. R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, CRC Press, 2021.
6. Y. Li and J. S. Racine, *Cross-validated local linear semiparametric regression for mixed data*, Journal of Applied Econometrics, vol. 35, no. 5, pp. 565–586, 2020.

7. X. Luo, J. Zhang, and L. Wang, *Semiparametric regression for left truncated and interval censored medical data using kernel smoothing*, Statistical Methods in Medical Research, vol. 30, no. 9, pp. 1951–1967, 2021.
8. S. N. Wood, *Inference and computation with generalized additive models and their extensions*, TEST, vol. 29, no. 2, pp. 307–339, 2020.
9. G. Hesamian and M. G. Akbari, *Fuzzy regression models: A review and new perspectives*, Soft Computing, vol. 25, pp. 13429–13452, 2021.
10. A. Gasparrini, F. Scheipl, B. Armstrong, and M. G. Kenward, *A penalized framework for distributed lag non-linear models*, Biometrics, vol. 78, no. 3, pp. 938–948, 2022.
11. S. M. Mousavi and M. Keshtkaran, *A fuzzy semiparametric regression model using α-cut decomposition for mixed crisp–fuzzy data*, Soft Computing, vol. 26, pp. 12345–12360, 2022.
12. A. Khosravi and R. Fattahi, *A fuzzy mathematical programming approach to semiparametric regression models with uncertain data*, Expert Systems with Applications, vol. 225, 120012, 2023.
13. A. Khosravi and R. Fattahi, *A fuzzy optimization framework for semiparametric regression models with uncertain inputs*, Expert Systems with Applications, vol. 225, 120012, 2023.
14. L. A. Zadeh, *A note on Z-numbers*, Information Sciences, vol. 181, no. 14, pp. 2923–2932, 2011.
15. M. Mohammadi, S. M. Taheri, and G. Hesamian, *Regression analysis under Z-number uncertainty: Concepts and modeling approaches*, Soft Computing, vol. 26, pp. 12079–12093, 2022.
16. R. A. Aliev and O. H. Huseynov, *Z-number-based fuzzy regression models: Theory and applications*, Fuzzy Sets and Systems, vol. 415, pp. 1–20, 2022.
17. S. Ezadi and T. Allahviranloo, *Numerical solution of linear regression based on Z-numbers by improved neural network*, Intelligent Automation & Soft Computing, vol. 23, no. 1, pp. 1–11, 2017.
18. A. Azadeh, M. Saberi, and A. Gitiforouz, *A new fuzzy regression framework for modeling uncertain systems*, Applied Soft Computing, vol. 90, 106154, 2020.
19. R. A. Aliev, O. Huseynov, and A. Alizadeh, *Z-number based decision making and modeling under reliability information*, Applied Soft Computing, vol. 104, 107256, 2021.
20. M. A. Abdulhussein, M. Niaparast, and S. Ezadi, *Fuzzy semi-parametric regression based on Z numbers*, Transactions on Fuzzy Sets and Systems, advance online publication, 2026.
21. L. M. Zeinalova, O. H. Huseynov, and P. Sharghi, *A Z-number valued regression model and its application*, Intelligent Automation & Soft Computing, vol. 23, no. 1, pp. 1–5, 2017.
22. J. M. Rodríguez Poó and A. Soberón, *Semiparametric estimation of partially linear models using kernel regression methods*, Journal of Applied Econometrics, vol. 36, no. 4, pp. 487–505, 2021.
23. C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, *Algorithm 778: L-BFGS-B: Fortran subroutines for large scale bound constrained optimization*, ACM Transactions on Mathematical Software, vol. 23, no. 4, pp. 550–560, 1997.
24. N. Rashad, N. Hammood, and Z. Algamal, *Linearized shrinkage estimator for semiparametric regression model*, Statistics, Optimization & Information Computing, 2026.
25. O. Khazaei, M. Ganjali, and M. Khazaei, *A Bayesian semi-parametric quantile regression approach for joint modeling of longitudinal ordinal and continuous responses*, Statistics, Optimization & Information Computing, vol. 11, no. 2, pp. 445–464, 2023.
26. R. A. Aliev and A. V. Alizadeh, *Algebraic properties of Z-numbers under multiplicative arithmetic operations*, in Advances in Intelligent Systems and Computing, vol. 896, pp. 33–41, Springer, 2019.