International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 2
ISSN: 2088-5334

Using Opposition Golden Jackal Optimization Algorithm (OGJO) in Improving Some Kernel Semiparametric Models: A Comparative Study

Dalia Mahfood Abdulhadi [a], Husam Abdulrazzak Rasheed [a,*]

[a] *Faculty of Administration and Economics, Mustansiriyah University, Baghdad, Iraq*
Corresponding author: [*]husamstat@uomustansiriyah.edu.iq

***Abstract*—Recent research and studies show widespread interest in semiparametric regression model analysis, which combines parametric and nonparametric components. This interest is because it gives accurate and effective statistical model estimates. This paper proposes to improve estimates of semiparametric regression models using opposition-based learning technology on the Golden Jackal Optimization (GJO) algorithm to increase the accuracy of these models, accelerate convergence, and expand the exploration area. The effectiveness of using this algorithm was evaluated by comparing it with the original algorithm before optimization and the most commonly used methods for estimating the model statistically, such as CV and GCV. Using simulation, the results showed that the improvement in the OBL-GJO algorithm in terms of accuracy and convergence speed outperformed the original algorithm and traditional methods by a large margin in calculating the simulation results of the kernel semiparametric regression models. We strongly advocate for applying the GJO algorithm across various domains within machine learning, particularly in the realms of deep learning and reinforcement learning. Furthermore, we have employed enhanced and evolved algorithms to optimize semiparametric regression models effectively. To address the challenges encountered by any algorithm operating within a vast search landscape, we suggest an in-depth exploration of optimization techniques and integrating diverse algorithms, which could lead to more robust and efficient solutions.**

Keywords—Semiparametric regression model; GJO; OBL-GJO; kernel model.

Manuscript received 15 Feb. 2024; revised 29 Apr. 2024; accepted 12 Dec. 2024. Date of publication 30 Apr. 2025.

I. INTRODUCTION

The modeling phenomenon is one of the main objectives in studying various problems or phenomena, including economic, engineering, medical, and other areas. It involves finding a mathematical model that serves as a primary means to understand these phenomena' patterns and behaviors, and analyze their data [1]. In some cases, particularly with economic phenomena, the model that describes the situation does not meet the assumptions of a parametric model, leading to inconsistencies. Consequently, the functional form of the phenomenon remains undetermined without adequate knowledge about the subject under study [2]. There are numerous reasons, such as conducting an experiment for the first time, being unable to ascertain the causal relationships between variables, or failing to fully explain the relationship between the response variable and the explanatory variables. Non-parametric models, while offering high flexibility, face challenges associated with the curse of dimensionality [3]. The curse of dimensionality is a common issue in nonparametric models, which becomes apparent as the number of dimensions (variables) in the data significantly increases, resulting in difficulties in analysis and interpretation [4]. Estimates in nonparametric models rely on the available data without imposing strong assumptions about the relationship between variables, making them sensitive to rising dimensions. This sensitivity leads to various problems in analysis and modeling [5].

Here, semiparametric regression models emerged, combining the characteristics of parametric and nonparametric regression models. This model is considered one of the hybrid models, as it integrates both types. It encompasses all the positive features of the previous two models. In this paper, we estimated the regression function of the kernel semiparametric model using two methods (the smoothing matrix method and a different approach) [6].

The most common statistical methods used to evaluate model performance and enhance accuracy include the CV and GCV methods. Additionally, it incorporates the most

effective algorithms currently applied in optimization (the Golden Jackal Improvement algorithm) [7], [8]. It employs opposition learning technology, utilizing opposition-based learning within the Golden Jackal Improvement algorithm to improve the kernel parameter selection in the semi-parametric model [9]. Consequently, it addresses statistical problems and identifies optimal opportunities to enhance results, which is evident when comparing the simulation results and elucidating them in Economic Data (Money Supply) [10].

The most important references on parametric and nonparametric regression models have been included in [11], [12], as well as a detailed explanation of the estimators of those models. Another study by [13] has proposed a new way to find missing data (semiparametric data). AI is used to bandwidth in the kernel method and applied in simulation and real data. The results demonstrated that calculating random semiparametric regression is better than the current deterministic semiparametric regression regarding efficiency and effectiveness. The simulation results indicate that the estimator is superior to many designs, and this estimator is consistent and natural without convergence [14]. A cumulative incidence function model with missing and variable values has been proposed by [15]. The study has introduced a new class of generalized semiparametric regression models for CIF based on the missing variable distribution estimator and linear regression model. The study has used simulations to compare the results and then applied the models based on real data of patients with myelogenous leukemia [16].

A previous study by [17] has employed the semiparametric mixed effects model through longitudinal data analysis. The kernel approach was used to estimate the nonparametric function, and the weighted least squares method was utilized to estimate the regression coefficients. They discovered that the back-fitting procedure frequently produces additional bias and contrast. An iterative technique is used to estimate the variance function to enhance efficiency. The study employed the bat algorithm in a different space when determining the homogeneity coefficient for semiparametric regression models. The results have shown that the technique based on optimal values of the homogeneity coefficient was more accurate than other comparison methods, including CV and GCV [18].

Considering the revolution of improving efficiency and performance and developing data modeling and analysis concepts to reach the best and most accurate results. It improves quality and facilitates integration between systems to reduce redundancy and build on them in making future decisions for institutions and countries. Significantly, the economic and financial aspects need good results to be built upon [19]. Especially in the absence of complete knowledge of the behavior of these variables in those data, it is impossible to determine the causal or behavioral relationship that links all these variables. It leads to problems in choosing the appropriate model for the phenomenon. Estimates of inaccurate parameters, such as assuming a linear relationship between variables and neglecting the effect of nonlinear variables, are made. Complexity in processing and large time consumption results from the multiplicity of variables and the magnitude of data in an ample research space [20]. The above and other complications lead to apparent problems in the accuracy of description and prediction of future values. This is reflected in the analysis of financial statements when using traditional estimation methods.

II. Materials and Method

A. Semiparametric Regression Models

Semiparametric regression models have gained interest in advanced statistical analyses to obtain highly efficient estimators [21]. They combine rigorous parametric regression models with highly flexible parametric regression models. The partial linear regression model (PLRM) is a standard parametric regression model where the relationship between parametric and nonparametric parts is displayed [22]. Robinson [23] proposed the PLRM model. The general form of the semiparametric regression model is:

$$y_i = \sum_{j=1}^{p} \beta_j Z_{ij} + g(u_i) + \varepsilon_i \quad , i = 1,2,3 \dots, n \qquad (1)$$

whereas,

y: represents the response variable vector or dependent variable and is of the degree n × 1.

Z : represents a matrix of explanatory variables of the degree (n×(p+1)).

β : Represents the vector of unknown features to be estimated in degrees ((p + 1) ×1).

ε : is a vector of random errors of degree n × 1 and is independent with mean 0 and variance σ^2.

u_i Continuous variable that expresses the nonparametric variable in dimensions (nx1).

The model described in the previous equation in matrix form is:

$$y = Z\beta + g + \varepsilon \qquad (2)$$

where:

Zβ : The parametric part of the model under study.

g: An indefinite prime function of the variable (u) of degree (nx1), which represents the nonparametric part of the model [24].

B. Parameters of the Parametric Part

We estimated the parameters of the parametric model using the least squares method, the mathematical formula of which is as follows [25].

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j Z_{ij} + \varepsilon_i \quad , i = 1,2,3 \dots, n \qquad (3)$$

The multiple linear regression model is commonly used to study the effect of numerous explanatory variables on the response variable, and its mathematical formula. It consists of (p) explanatory variables, (p + 1) parameters, and (n) observations. Using the matrix method, the formula is:

$$Y = Z\beta + \varepsilon \qquad (4)$$

Its parameters are estimated using the least squares (OLS) method, and their formulas are:

$$\hat{\beta}_{OLS} = (Z^T Z)^{-1} Z^T y \qquad (5)$$

Then, we estimate the parameters of the nonparametric model using the smoothing matrix method [26].

We also utilized a semi-parametric regression model (Kernel smoother). The kernel semi-parametric regression model is one of the most essential methods to estimate the nonlinear component in semi-parametric models by

employing a kernel function (Kernel smoother) or, in other words, by altering the data representation [27].

$$\begin{aligned}\phi: X = (x_1, x_2, \ldots, x_N) &\rightarrow Z \\ &= \left(z_1 = \phi(x_1), z_2 \right. \\ &= \left. \phi(x_2), \ldots, z_N = \phi(x_N)\right)\end{aligned} \quad (6)$$

Two kernel functions were used :
Polynomial kernel functions

$$K(x, y) = [x^T y + 1]^m \quad (7)$$

Gaussian radial basis kernel functions

$$K(x, y) = \frac{exp(-\|x-y\|^2)}{2\sigma^2} \quad (8)$$

The semi-parametric regression model is defined as follows:
Polynomial:

$$Y = Z\beta + [x^T y + 1]^m + e \quad (9)$$

Gaussian radial basis:

$$Y = Z\beta + exp(-\|x - y\|^2/2\sigma^2) + e \quad (10)$$

C. Cross-Validation and Generalized Cross-Validation Method

It is a technique widely used in machine learning and statistics to evaluate the performance of a predictive model. It helps estimate how well the model performs on unobserved data [28], [29]. Below are the detailed rules of the cross-validation method:

1) Data partitioning: The data set is divided into two parts: a training set and a test set.

2) K-Fold Cross Validation: The data set is divided into subsets (segments) of approximately equal size K clusters.

3) Training and testing iterations: The cross-validation process consists of K iterations.

4) Performance metrics: Performance metrics such as accuracy or mean squared error (MSE) are calculated for each iteration.

5) K-Fold cross-validation frequency: Steps 3 and 4 are repeated K times, each time using a different group as the test set.

6) Evaluation: The performance metrics obtained in each iteration are averaged to find a single estimate of the model's performance.

7) Parameter setting: Cross-validation is also used to tune hyperparameters.

8) Final model: After cross-validation, the final model can be trained on the entire dataset using the best-performing hyperparameters and evaluated on an utterly unseen dataset to estimate its performance in real-world scenarios.

The detailed rules of the GCV Method are as follows [30]:

1) Split data: The dataset is divided into a training group and a test set.

2) Model training: The model is trained in the training group.

3) Performance evaluation: The test group evaluates the model's performance and calculates predictive errors.

4) Generalized Cross-Validation Calculation: The GCV value is calculated, which reflects the model's accuracy in predicting new data.

5) Evaluation: A GCV score provides an estimate of model performance. A low GCV score indicates better data fits with good generalization of invisible data.

6) Choose the optimal model: Choose the model that reduces the GCV score, as the lower value indicates the best model.

7) Parameter adjustment: Generalized Cross-Validation is also used to adjust superparameters.

8) Final model: After the Generalized cross-validation process, the final model can be trained over the entire dataset using the best-performing superparameters and evaluated on an utterly invisible dataset to estimate its performance in real-world scenarios.

D. Improving Kernel Semi-parametric Model by Opposition-based Learning in Golden Jackal Optimization Algorithm

To enhance the procedure for choosing the kernel parameter in the parametric model, this technique aims to employ opposition-based learning in the initial community formation of the Golden Jackal Improvement algorithm. Here are some of the most fundamental ideas that this algorithm addressed [31]:

- GJO algorithm, which stands for Golden Jackal Optimization.
- O-BL stands for Opposition-Based Learning.
- Make use of OBL in the OGJO Algorithm with the help of optimization

A cooperative hunting strategy the Golden Jackal (Canis aureus) used inspired GJO [32]. An Improvement Method for Golden Jackals (GJO). Prey search, confinement, and pounce are mathematically modeled and implemented in three basic steps. The suggested algorithm's performance on modular functions is evaluated in contrast to state-of-the-art heuristics. The first step in solving an optimization problem with GJO is to initialize a set of possible solutions. A vector of values encoding the parameters or variables of each potential solution is used. At the first stage of the algorithm, called the "exploration phase", candidate solutions are evaluated based on their efficiency or function goal. The most suitable solutions are retained, while the less suitable ones are eliminated. The exploration phase is followed by the "exploitation phase", where retained solutions are used to direct the search toward better solutions. This stage is inspired by the Golden Jackal's tactic of working together to hunt prey. At this point, the retained solutions act as mentors or leaders, pulling other candidate solutions towards them. GJO has been proven effective in solving a wide range of optimization issues, including those that are nonlinear and polynomial. It is characterized by its ability to adapt to various optimization problems, and its performance has been positively compared to other common optimization algorithms such as ant colony algorithms and particle swarm optimization [33]. It can provide an understanding of the cooperative hunting behavior and strategies of the golden jackal in nature. This algorithm simulates foraging behaviors, infiltration, and surrounding prey to capture, helping to update the jackal's location and

improve solutions. GJO represents one solution inspired by golden jackal hunting tactics that have applications in various fields, such as robotics and artificial intelligence. In the beginning, community members are randomly selected and prepared according to the minimum and upper limit of the problem according to equation (11):

$$Z_0 = L + rand * (U - L) \quad (11)$$

where L and U represent the lower and upper boundaries of the lookup area, and rand refers to a random number defined by the interval [0,1]. In the equation below, the prey matrix is produced at this stage, and two of its elements are selected as the fittest jackal pair.

$$prey = \begin{bmatrix} Z_{1,1} & Z_{1,2} \dots & Z_{1,d} \\ Z_{2,1} & Z_{2,2} \dots & Z_{2,\mathrm{d}} \\ \vdots & \vdots & \vdots \\ Z_{\mathrm{n},1} & Z_{n,2} \cdots & Z_{\mathrm{n,d}} \end{bmatrix} \quad (12)$$

where the elementary solution, denoted as Z_(i,j), is the j^th. With d representing the number of variables and i^th representing the number of prey, the total number of prey is denoted by n. where the placement of the prey serves as an indicator of the solution parameters. The next phase is discovery, or the hunt for food. The GJO exploration strategy is now being proposed. Jackals are naturally good hunters; yet they seldom hunt because their prey is difficult to catch and escape from. Once they have encircled their prey, they leap on top of it to devour it. The following is the mathematical expression of this hunting activity, including a male and female jackal [34].

$$Z_1(t) = Z_M(t) - E.\lceil Z_M(t) - rl * prey(t)\rceil \quad (13)$$

$$Z_2(t) = Z_{FM}(t) - E.\lceil Z_{FM}(t) - rl * prey(t)\rceil \quad (14)$$

The goal of rl is included in the formula. (1), (2), and (3). The goal is to generate irrational behavior during exploitation, emphasizing discovery and avoiding the readily available local optimal solution. Prey evasion energy is represented by E. Finally, the following is how the new Golden Jackal update's location is determined:

$$Z(t+1) = \frac{Z_1(t) + Z_2(t)}{2} \quad (15)$$

After the prey is located, the jackal pairs encircle or pounce on it, depending on the situation. At this point, the victim's ability to flee is compromised due to the jackal's harassment. They pounce on their target after encirclement and devour it. Mathematically, this hunting behavior looks like this:

$$Z_1(t) = Z_M(t) - E.\lceil rl.Z_M(t) - prey(t)\rceil \quad (16)$$

$$Z_2(t) = Z_{FM}(t) - E.\lceil rl.Z_{FM}(t) - prey(t)\rceil \quad (17)$$

E. Opposition-Based Learning (O-BL)

When applied to complicated situations, especially those involving research and improvement, the concept of O-BL was presented as a new technology that might speed up the convergence of evolutionary approaches. The sudden and dramatic changes in social progress serve as inspiration. Crucial to this iteration's operation is that, similar to initial population preparation, each particle or potential optimization problem solution can be constructed in one of two ways: by incorporating data collected throughout algorithm development or by making basic random guesses [35]. In the second stage, increasing the convergence speed can likely be achieved by initially and iteratively producing an inverse point for each potential solution. According to O-BL, a more transparent method of determining the optimal value may be to investigate both the forward and backward directions of the initial candidate solution. Consequently, the essential concept is the inverse of a number; for instance, given a set of solutions to a problem, we can find the inverse of each of those solutions. This can hasten the search for ideal solutions and enhance performance. This is where the inverse of any given real value Z ∈ [L,U] may be found using the following.

$$\bar{Z} = L + U - Z \quad (18)$$

where the real position vector Z∈R^n is opposed to the imaginary position vector $\bar{Z}$∈R^n. Another step-in optimization is comparing the fitness functions of the two solutions; the better ones are saved, and the worst ones are removed. As an illustration, Z is preserved if F(Z) is smaller than F($\bar{Z}$) (to zoom out), while $\bar{Z}$ is preserved if F(Z) is larger than F($\bar{Z}$).

F. Combination of OBL and GJO Algorithms

In this section, the suggested OGJO is shown as an amalgam of GJO algorithm features and OBL technology that improves the accuracy of the ideal answer by expanding the scope of study. Due to its many flaws, OBL improves the GJO approach. These flaws include slow convergence, a tendency to settle for adjacent answers, a lack of investigation of the study region, and an inefficient use of time. Avoiding these problems is the goal of the suggested method, which covers the study area by considering the two possible outcomes of the computed point, which entails contradictory values. This enhancement enhances the probability of using optimal solutions in half the time. There are two phases to executing the suggested plan of action.

G. Optimization in Kernel Semiparametric Model

The superparameters associated with the type of kernel have a direct impact on improving the solution of the target function. The performance of the kernel function depends heavily on the choice of these superparameters, but no mathematical methods are available to determine their exact values. Thus, the research on the parametric regression model of the nucleus depends heavily on selecting these superparameters. While the CV and GCV methods, as these are the most common techniques used in literature to determine these parameters, are time-consuming. Optimized algorithms are used in computing, resulting in more appropriate performance than any algorithm alone [36]. One such algorithm is OGJO (see the diagram in Fig. 1), which can overcome local optimal solutions and share features into a compromise between exploration and exploitation, delaying the beginning of each algorithm's convergence, and preventing falling into the local optimal level. To improve the performance characteristics of the proposed algorithm and obtain faster convergence to escape from local optimization, which enhances the exploration and exploitation of the GJO algorithm, making it more effective. This efficiently helps to find the most essential superparameter values related to the semiparametric kernel model with high prediction performance. A flowchart to optimize the algorithm used for the kernel semiparametric model is displayed (see Fig. 2) as follows:

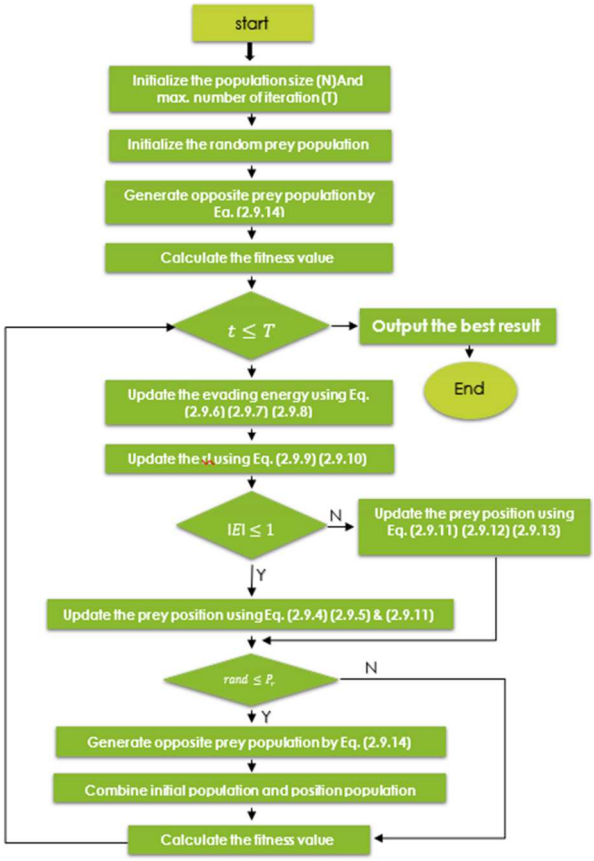

Fig. 1 OGJO algorithm diagram

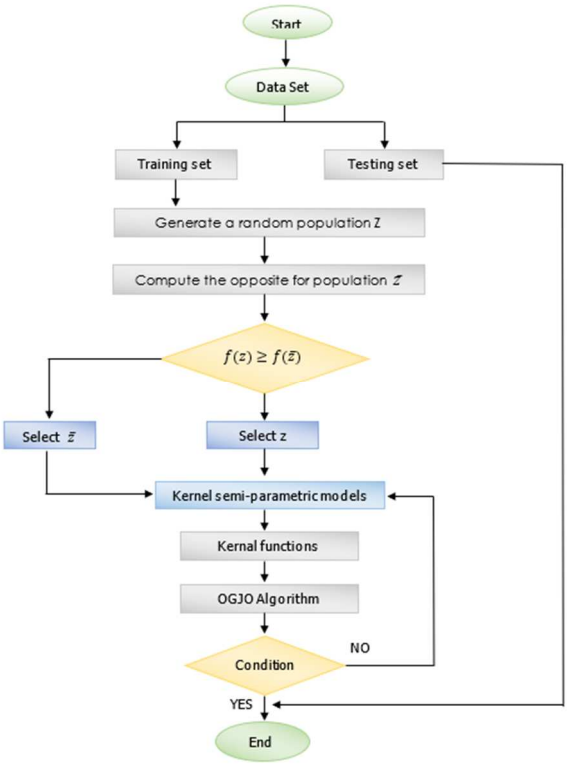

Fig. 2 Semi-parametric kernel model using the OGJO

- Twenty-five Golden Jackals were allocated, and a maximum of t = 500 repetitions were planned.
- Golden Jackal exhibitions were developed through the concept of opposition learning.
- For each of the two groups, in the first iteration, the best and second (male and female Golden Jackal) are calculated between the set of elements and the corresponding elements.
- Male and female Golden Jackals are placed in their positions randomly. The superparameter P (the parameter of the kernel function) is represented by Golden Jackal position. Golden Jackal starts randomly and is selected from the normal distribution within the interval.
- Calculate $\bar{Z}$ for each superparameter using equation (18).
- The fitness function is defined as follows:

$$fitness = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \qquad (19)$$

- Locations are updated using equations (13), (14).
- 8. Repeat steps 3,4 and 5 to get to t_{max} .

III. Results and Discussion

This is in line with the study's objective. Both simulation experiments and the application of models using near-real data provide evidence of the validity of theoretical assumptions. These results emphasize the importance of choosing the appropriate estimation method when dealing with real data.

A. Simulation Result

Many experiments are designed and simulated using one of the statistical programming languages R(Ver.4.0.3) to generate and process data statistically, The effect of sample size was taken into account on the estimators under study, where three different sizes of samples n (50,100,200) were used, and to reduce the variation in the generation of simulation data, the experiment was repeated (1000) times, Comparisons were made for the different methods used CV and GCV with Golden Jackal optimization algorithm and the proposed algorithm OGJO using the Gaussian Radial basis function and the Polynomial function as the kernel functions, with four different models being studied.

1) First model: This model was generated according to the following regression equation parameters: $\beta = (-1.5,3,2,-5,4)^T$ The values of the variable (u_i) of the nonparametric component is generated according to the uniform distribution and within the interval $u \in [0,1.5]$ which can be used to find the nonparametric function according to the following equation:

$$g(u_i) = \exp\left(\sin(5u_i) + \sqrt[4]{u_i}\right)$$

In addition to the distribution of errors $\varepsilon_i \sim N_n(0, \sigma_v^2)$, $\sigma^2 = 0.64$, $v_{ij} = \exp(-3|i-j|)$

2) Second model: In this model, we use the regression equation parameters.

$\beta = (-1,-1,2,3,-5,4)^T$ The values of the variable (u_i) of the nonparametric component is generated according to the uniform distribution, within the interval $u \in [-9,9]$, which can be used to find the nonparametric function according to the following formula:

$$g(u) = \frac{1}{5}\left[\begin{matrix}\emptyset(u;-7,1.44) + \emptyset(u;-3.5,1) + \emptyset(u;0,0.64) \\ +\emptyset(u;3.5,0.36) + \emptyset(u;7,0.16)\end{matrix}\right]$$

As for the distribution of errors, $\varepsilon_i \sim N_n(0, \sigma_v^2)$, $\sigma^2 = 4$, $v_{ij} = (\frac{1}{n})^{|i-j|}$

3) Third model: This model was generated according to the following equation parameters]: $\beta = (1.5,2,3,-5,4)^T$ Also, the values of the variable (u_i) of the nonparametric component was generated according to the formula, $\boldsymbol{u_i} = \frac{(\boldsymbol{i-0.5})}{\boldsymbol{n}}$,within the interval $u \in [-9,9]$ which can be used to find the nonparametric function according to the following formula:

$$g(u_i) = \sqrt{u_i(1-u_i)}\ \sin\left(\frac{2.1\pi}{u_i+0.05}\right)$$

As for the distribution of errors, it is as follows $\varepsilon_i \sim N_n(0, \sigma^2)$, $\sigma^2 = 0.01$

4) Fourth model: This model was generated according to the following equation parameters.

$$\beta = (1,-1,2,3-5,4)^T$$

Also, the values of the variable (u_i) of the nonparametric component were generated according to the uniform distribution, within the interval $u \in [0,1]$,which can be used to find the nonparametric function according to the following formula:

$$g(u_i) = \sin(2u_i)\cos(5u_i)$$

As for the distribution of errors, it is as follows $\varepsilon_i \sim Nn(0, \sigma_v^2)$, $\sigma^2 = 0.09$, $v_{ij} = (-9exp|i-j|)$

B. Application of Mean Absolute Percentage Error

The comparison between the above estimation methods and the statement of the best estimator is made using MAPE (Mean Absolute Percentage Error), in the following formula:

$$MAPE = 100\frac{1}{n}\sum_{i=1}^{n}\frac{y_i-\hat{y}}{y_i} \qquad (20)$$

The results were shown using the Gaussian Radial basis function with four models.

TABLE I
MAPE VALUE FOR THE USED METHODS OF THE FIRST MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000183	0.000162	0.000152	0.000132
GCV	0.000180	0.000159	0.000149	0.000129
GJO	0.000155	0.000133	0.000124	0.000103
OGJO	0.000148	0.000127	0.000117	0.000097

TABLE II
MAPE VALUE FOR THE USED METHODS OF THE SECOND MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000280	0.000258	0.000248	0.000228
GCV	0.000271	0.000250	0.000239	0.000219
GJO	0.000245	0.000224	0.000214	0.000193
OGJO	0.000243	0.000222	0.000212	0.000191

TABLE III
MAPE VALUE FOR THE USED METHODS OF THE THIRD MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000382	0.000361	0.000351	0.000331
GCV	0.000373	0.000352	0.000342	0.000322
GJO	0.000347	0.000327	0.000317	0.000296
OGJO	0.000345	0.000324	0.000314	0.000294

TABLE IV
MAPE VALUE FOR THE USED METHODS OF THE FOURTH MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000369	0.000348	0.000338	0.000317
GCV	0.000360	0.000339	0.000329	0.000308
GJO	0.000334	0.000313	0.000303	0.000283
OGJO	0.000332	0.000311	0.000301	0.000281

C. Application of Polynomial Kernel Semiparametric Model

And also, other results shown by using the Polynomial kernel semiparametric function with the same four models.

TABLE V
MAPE VALUE FOR THE USED METHODS OF THE FIRST MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000197	0.000176	0.000166	0.000145
GCV	0.000193	0.000172	0.000162	0.000142
GJO	0.000168	0.000147	0.000137	0.000116
OGJO	0.000161	0.000140	0.000130	0.000110

TABLE VI
MAPE VALUE FOR THE USED METHODS OF THE SECOND MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000293	0.000272	0.000262	0.000241
GCV	0.000284	0.000263	0.000253	0.000232
GJO	0.000259	0.000237	0.000227	0.000207
OGJO	0.000256	0.000235	0.000225	0.000205

TABLE VII
MAPE VALUE FOR THE USED METHODS OF THE THIRD MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000395	0.000374	0.000364	0.000344
GCV	0.000386	0.000365	0.000355	0.000335
GJO	0.000360	0.000340	0.000330	0.000310
OGJO	0.000359	0.000338	0.000328	0.000307

TABLE VIII
MAPE VALUE FOR THE USED METHODS OF THE FOURTH MODEL

Method	n=30	n=100	n=200	n=500
CV	0.000382	0.000361	0.000351	0.000331
GCV	0.000373	0.000352	0.000342	0.000322
GJO	0.000347	0.000327	0.000317	0.000296
OGJO	0.000345	0.000324	0.000314	0.000294

The results showed that the OGJO method is superior to the other traditional methods (GCV, CV) and the algorithm before improving GJO optimization in the MAPE. By the sample size, we notice that the value of MAPE decreases as the sample size increases, while the preference remains for our OGJO proposed algorithm in all tables.

IV. CONCLUSION

In conclusion, the proposed Opposition-Based Learning in Golden Jackal Optimization Algorithm (OGJO) works better than the Golden Jackal Algorithm. The optimization algorithm (OGJO) reduces the calculation time required for estimation compared to other methods, such as CV and GCV. OGJO explores the search space more effectively than other CV and GCV measurement methods. The proposed algorithm has the potential to enhance the applicability of parametric kernel models in various applications, including finance, economics, engineering, and medicine. The GCV method achieved better results in estimating the smoothing matrix of parameters compared with the CV method and for all models adopted in the simulation. We recommend applying the

algorithm (GJO) in other areas of machine learning, such as deep learning and reinforcement learning. Also, improved and developed algorithms were used to optimize semiparametric regression models. To overcome the problems experienced by any algorithm in an extensive search area, we recommend researching the optimization and integration of algorithms.

ACKNOWLEDGMENT

No grant funded this research.

REFERENCES

[1] C. Yang, Y. Gao, W. Guo, Y. Yang, P. Xu, and M. S. Alqahtani, "High-velocity impact behaviour of curved GFRP composites for rail vehicles: Experimental and numerical study," *Polym. Test.*, vol. 116, p. 107774, Dec. 2022, doi: 10.1016/j.polymertesting.2022.107774.

[2] F. Chen, Y. Peng, X. Chen, K. Wang, Z. Liu, and C. Chen, "Investigation of the Ballistic Performance of GFRP Laminate under 150 m/s High-Velocity Impact: Simulation and Experiment," *Polymers (Basel).*, vol. 13, no. 4, p. 604, Feb. 2021, doi:10.3390/polym13040604.

[3] E. Admasu, S. J. Thirgood, A. Bekele, and M. Karen Laurenson, "Spatial ecology of golden jackal in farmland in the Ethiopian Highlands," *Afr. J. Ecol.*, vol. 42, no. 2, pp. 144–152, Jun. 2004, doi:10.1111/j.1365-2028.2004.00497.x.

[4] E. Akdeniz, F. Akdeniz, and M. Roozbeh, "A new difference-based weighted mixed Liu estimator in partially linear models," *Statistics (Ber).*, vol. 52, no. 6, pp. 1309–1327, Nov. 2018, doi:10.1080/02331888.2018.1511715.

[5] M. Amini and M. Roozbeh, "Optimal partial ridge estimation in restricted semiparametric regression models," *J. Multivar. Anal.*, vol. 136, pp. 26–40, Apr. 2015, doi: 10.1016/j.jmva.2015.01.005.

[6] S. Chen and X. Zhou, "Semiparametric estimation of a truncated regression model," *J. Econom.*, vol. 167, no. 2, pp. 297–304, Apr. 2012, doi: 10.1016/j.jeconom.2011.09.016.

[7] S. H. Chiou, C. Huang, G. Xu, and J. Yan, "Semiparametric Regression Analysis of Panel Count Data: A Practical Review," *Int. Stat. Rev.*, vol. 87, no. 1, pp. 24–43, Apr. 2019, doi:10.1111/insr.12271.

[8] N. Chopra and M. Mohsin Ansari, "Golden jackal optimization: A novel nature-inspired optimizer for engineering applications," *Expert Syst. Appl.*, vol. 198, p. 116924, Jul. 2022, doi:10.1016/j.eswa.2022.116924.

[9] E. A. Duran and F. Akdeniz, "New difference-based estimator of parameters in semiparametric regression models," *J. Stat. Comput. Simul.*, vol. 83, no. 5, pp. 810–824, May 2013, doi:10.1080/00949655.2011.638633.

[10] E. Ghysels and H. Qian, "Estimating MIDAS regressions via OLS with polynomial parameter profiling," *Econom. Stat.*, vol. 9, pp. 1–16, Jan. 2019, doi: 10.1016/j.ecosta.2018.02.001.

[11] R. Guha, M. Ghosh, A. Chakrabarti, R. Sarkar, and S. Mirjalili, "Introducing clustering based population in Binary Gravitational Search Algorithm for Feature Selection," *Appl. Soft Comput.*, vol. 93, p. 106341, Aug. 2020, doi: 10.1016/j.asoc.2020.106341.

[12] W. Härdle, *Nonparametric and semiparametric models*. Springer Science & Business Media, 2004.

[13] U. Lee, Y. Sun, T. H. Scheike, and P. B. Gilbert, "Analysis of generalized semiparametric regression models for cumulative incidence functions with missing covariates," *Comput. Stat. Data Anal.*, vol. 122, pp. 59–79, Jun. 2018, doi: 10.1016/j.csda.2018.01.003.

[14] C. A. Micchelli, M. Pontil, and P. Bartlett, "Learning the Kernel Function via Regularization.," *J. Mach. Learn. Res.*, vol. 6, no. 7, 2005.

[15] Y. Qin, S. Zhang, X. Zhu, J. Zhang, and C. Zhang, "Semi-parametric optimization for missing data imputation," *Appl. Intell.*, vol. 27, no. 1, pp. 79–88, Jun. 2007, doi: 10.1007/s10489-006-0032-0.

[16] P. M. Robinson, "Large-sample inference for nonparametric regression with dependent errors," *Ann. Stat.*, vol. 25, no. 5, pp. 2054–2083, 1997.

[17] M. Roozbeh, "Shrinkage ridge estimators in semiparametric regression models," *J. Multivar. Anal.*, vol. 136, pp. 56–74, Apr. 2015, doi:10.1016/j.jmva.2015.01.002.

[18] M. Roozbeh, G. Hesamian, and M. G. Akbari, "Ridge estimation in semi-parametric regression models under the stochastic restriction and correlated elliptically contoured errors," *J. Comput. Appl. Math.*, vol. 378, p. 112940, Nov. 2020, doi: 10.1016/j.cam.2020.112940.

[19] C. Schramm, S. Jacquemont, K. Oualkacha, A. Labbe, and C. M. T. Greenwood, "KSPM: A Package For Kernel Semi-Parametric Models," 2020.

[20] I. O. Bachi Al-Fahad, H. K. Sharaf, L. N. Bachache, and N. K. Bachache, "Identifying the Mechanism of the Fatigue Behavior of the Composite Shaft Subjected to Variable Load," *Eastern-European J. Enterp. Technol.*, vol. 123, no. 7, 2023.

[21] L. T. Mouhmmd, M. A. Rahima, A. M. Mohammed, H. F. Hasan, A. S. Alwan, and H. K. Sharaf, "The effect of firm type on the relationship between accounting quality and trade credit in listed firms," *Corp. Bus. Strateg. Rev.*, vol. 4, no. 2, pp. 175–183, Apr. 2023, doi:10.22495/cbsrv4i2art16.

[22] H. A. Saleh, A. R. Ali, A. N. S. Almshabbak, H. K. Sharaf, H. F. Hasan, and A. S. Alwan, "The impact of auditor-client range on audit quality and timely auditor report," *Corp. Bus. Strateg. Rev.*, vol. 5, no. 1, special Issue, pp. 329–335, Feb. 2024, doi: 10.22495/cbsrv5i1siart7.

[23] P. M. Robinson, "Root-N-Consistent Semiparametric Regression," *Econometrica*, vol. 56, no. 4, p. 931, Jul. 1988, doi: 10.2307/1912705.

[24] E. Rouza and F. Arifin, "Essential Advances in Soil-Transmitted Helminth Detection Using Machine Learning and Deep Learning: A Systematic Review.," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 6, 2024.

[25] B. A. Sadkhan, E. J. Yousif, A. T. Shomran, E. K. Hussein, and H. K. Sharaf, "Investigation of the Impact Response of Plain Weave E-Glass Composite Structure Based on the EN ISO 178 Standard," *J. Adv. Res. Appl. Mech.*, vol. 117, no. 1, pp. 118–127, 2024.

[26] I. O. Bachi Al-Fahad, A. D. Hassan, B. M. Faisal, and H. K. Sharaf, "Identification of Regularities in the Behavior of a Glass Fiber-Reinforced Polyester Composite of the Impact Test Based on ASTM D256 Standard," *Eastern-European J. Enterp. Technol.*, vol. 124, no. 7, 2023.

[27] J. H. Connell and S. Mahadevan, Eds., Robot Learning. Boston, MA, USA: Kluwer, 1993/1997. [Book Review]. Robotica, vol. 17, no. 2, pp. 229-235, 1999, doi: 10.1017/S0263574799271172.

[28] M. Taavoni and M. Arashi, "Kernel estimation in semiparametric mixed effect longitudinal modeling," *Stat. Pap.*, vol. 62, no. 3, pp. 1095–1116, Jun. 2021, doi: 10.1007/s00362-019-01125-8.

[29] T. Zhang, "Semiparametric model building for regression models with time-varying parameters," *J. Econom.*, vol. 187, no. 1, pp. 189–200, Jul. 2015, doi: 10.1016/j.jeconom.2015.02.021.

[30] S. Yang, J. Wang, X. Cheng, J. Wu, and J. Liu, "Quality design based on kernel trick and Bayesian semiparametric model for multi-response processes with complex correlations," *Int. J. Prod. Res.*, vol. 62, no. 12, pp. 4407–4426, Jun. 2024, doi: 10.1080/00207543.2023.2262065.

[31] I. Pantelić, N. Bogdanović, S. Breck, and D. Ćirović, "Temporal trends of home range sizes and movement patterns of peri-urban golden jackals in Belgrade, Serbia," *Hystrix, Ital. J. Mammal.*, 2024.

[32] S. K. Venkatachary, A. Alagappan, and L. J. B. Andrews, "Cybersecurity challenges in energy sector (virtual power plants) - can edge computing principles be applied to enhance security?," *Energy Informatics*, vol. 4, no. 1, p. 5, Dec. 2021, doi: 10.1186/s42162-021-00139-7.

[33] I. O. B. Al-Fahad and H. K. Sharaf, "Investigation of the Effect of Heat Transfer during Friction Stir Welding (FSW) of AZ80A Mg Alloy Plates using a Pin Tool by Conducting Finite Elements Analysis," *J. Adv. Res. Fluid Mech. Therm. Sci.*, vol. 117, no. 1, pp. 98–108, 2024.

[34] M. Montaseri, M. Rezaei, A. Khayati, S. Mostafaei, and M. Taheri, "Survival parametric modeling for patients with heart failure based on Kernel learning," *BMC Med. Res. Methodol.*, vol. 25, no. 1, p. 7, Jan. 2025, doi: 10.1186/s12874-024-02455-4.

[35] X. Zhao, Y. Rong, J. Lin, M. Tian, and J. Liang, "Double sparsity garrotized kernel machine in high-dimensional partially linear model," *Commun. Stat. - Simul. Comput.*, pp. 1–16, Mar. 2024, doi:10.1080/03610918.2024.2329244.

[36] H. D. Salman, A. Q. Saeed, E. K. Hussein, H. K. Sharaf, T. Santos, and C. Santos, "Comparative Study on the Creep Behavior of a Hybrid Composite Material Doped with Nano Material," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 6, pp. 1951–1956, Dec. 2024, doi:10.18517/ijaseit.14.6.15958.