Research Article

A Robust Approach to CCRM Interval Regression considering Interval Coincidence Degree

Wang Yu and Yan Shilin

School of Airport Engineering and Transportation Management, Civil Aviation Flight University of China, Guanghan 618307, Sichuan, China

Correspondence should be addressed to Wang Yu; wangyu2001.111@163.com and Yan Shilin; 642874454@qq.com

Received 19 July 2021; Accepted 15 September 2021; Published 4 October 2021

Academic Editor: Ishfaq Ahmad

Copyright © 2021 Wang Yu and Yan Shilin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional CCRMs (Constrained Center-and-Range Methods) in solving the problem of interval regression could hardly make tradeoffs between the overall fitting accuracy and the coincidence degree between the observed and predicted intervals and could also hardly reduce the number of disjoint elements between the observed and predicted intervals, as well as raise the average ratio of all predicted intervals contained within their observed intervals. This paper constructed a nonlinear regression model based on center-and-range method, in which the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted interval was incorporated into the traditional CCRM model’s objective. This novel nonlinear programming model was proven to be a convex one that satisfied K-T condition. Monte Carlo simulation shows that the model is degenerated to the compared CCRM+ model as the objective only contains the minimization of the overall fitting accuracy for both center and range sample series. In this situation, it could obtain a better solution than the use of the compared CCRM model. In addition, when the proposed model only takes into account the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted interval, the model shows a better performance than the CCRM+ model in terms of the average ratio of all predicted intervals contained within their observed intervals, as well as the average number of forecasts with 0% accuracy.

1. Introduction

Studying the correlation between independent and dependent variables is one of the important tasks in data analysis, pattern recognition, artificial intelligence, and other related fields [1, 2]. Regression analysis is one of the main methods in determining the relation between the independent and dependent variables. Interval data are one of the symbolic data types, mainly formed by using technologies of both data-package and dimension-reduction to big data. Therefore, interval data regression refers to analyzing the relation between the interval independent and dependent variables [3–5], which is widely used in many industries, such as finance, engineering, medicine, transportation, etc.

Midpoint-Radius (called MR for simplicity) and Endpoint (called EP for simplicity) are two classic expression forms of interval data [6]. In the MR method, the interval data are converted into two types of point data series (namely, midpoint and radius series), such that the coefficients of two models are regressed, respectively. The EP method typically uses two types of point data series formed by the left and right endpoints of interval data, to carry out regression analysis. The min-max method is one of the classic EP methods in which the left and right endpoint data series of interval data are considered as two independent series, respectively [7]. A typical min-max method uses a convex combination of the left and right endpoints of interval independent variables to express the interval data, and finally to utilize Ordinary Least Squares (called as OLS for simplicity) method to obtain the regression coefficients [8]. Compared to the EP method, the MR method uses the midpoint series to represent the changing trend of interval...
data and utilizes the radius series to represent the uncertainty of interval data. Therefore, this kind of an expression is more consistent with the reality in industries.

To the best of our knowledge, the first study in interval data regression was released by Billard and Diday [9], in which the midpoint series of the interval data was used to estimate the regression coefficients. However, negligible complete information about the interval data could be captured. To solve this problem, the interval data series was converted into two independent data series, namely, midpoint and radius, respectively. Wei et al. [11] concluded that the midpoint and radius of interval data in actual situation are two types of data series with strong correlation, so the Seemingly Unrelated Regression Method (called SUR for simplicity) was developed to estimate the regression coefficients. Zhong et al. [12] were focused on the bias-corrected and heteroscedasticity-adjusted modeling technique by imposing order constraint to the endpoints of the response interval and the weighted linear least squares with estimated covariance matrix. Chacon and Rodriguez [13] presented a new approach to fit regression model for the symbolic interval-valued variables, which was essentially an improved and extended method for the center method proposed by Billard and Diday [9]. Lim [14] constructed a nonparametric interval regression model with an unknown smoothing function, and a backfitting algorithm was used to estimate the regression coefficients. However, in all abovementioned center-and-range methods (called CRM methods for simplicity), few considerations were given to the fact that the value of each right predicted endpoint must be greater than the value of the corresponding left predicted endpoint. Otherwise, the result was obviously inconsistent with the reality. To solve it, Lima Neto and de Carvalho [15] attempted to restrict the regression coefficient of the radius data series to be nonnegative during the fitting process, and formed a constrained linear regression model. This set of constraints made all predicted radius values also to be nonnegative, so as to ensure the reasonableness of the predicted intervals. Huiwen et al. [16] proposed a Complete Information Method (called CIM for simplicity) to ensure the reasonableness for each predicted interval. Guo and Hao [17] constructed a constrained center-and-range joint model to avoid the negative values of the predicted dependent interval variables. Other related studies were also attempted to solve such kind of an inconsistent problem [18–21]. However, when all abovementioned CCRM methods faced large fluctuations in midpoint data series and small fluctuations in radius data series, a poor predicted effect would appear due to a large number of predicted and observed intervals without any coincidences. Therefore, recent concentrations were paid to determine how to construct new constraints [22–24] to force this coincidence between the observed and predicted intervals. However, Monte Carlo simulation proposed by Guo et al. [22] discovered that this kind of a forced coincidence was actually at the expense of overall fitting accuracy for interval data sample series. This phenomenon became even worse when several abnormal intervals exist in data series. In reality, most decision-makers had little attention on appearing of a large number of disjoint elements between the observed intervals and predicted intervals, so overall fitting accuracy is low. Because neither of them could capture a satisfying information from the interval data. Therefore, determining how to present a method that can flexibly adjust such kind of trade-offs between the coincidence degree and the overall fitting accuracy is of great importance in solving these existing problems.

Therefore, the main contributions of this paper lie in the fact that: (1) a mathematical expression to this kind of a coincidence for the predicted and observed intervals is presented; (2) a nonlinear programming model with constraints that (a) each of the predicted and observed intervals must be intersected and (b) the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted interval is proposed; (3) the presented model is proven to be a convex programming model that satisfies K–T condition to solve; and (4) Monte Carlo simulation and empirical example are both used to verify the model’s feasibility.

The remainder of this paper is organized as follows. In Section 2, main CCRM methods are introduced. Section 3 establishes a novel CCRM regression method with a consideration of the interval coincidence degree, and then it is proven to be a convex programming model. Section 4 verifies the feasibility using Monte Carlo simulation on statistics. Section 5 is an empirical example with the data from airline industry.

2. Main CCRM Methods

Let \( E = \{e_1, e_2, \ldots, e_n\} \) be a \( p + 1 \) dimensional data set with a sample of \( n \). For each \( e_i \in E \), it can be set as \( e_i = (x_{i1}, x_{i2}, \ldots, x_{ip}, y_i), i \in \{1, 2, \ldots, n\} \), where \( x_{ij} (\forall j \in \{1, 2, \ldots, p\}) \) is an observation of the \( j \)th interval independent variable \( X_j \), where \( x_{ij} \in \{[\bar{x}_{ij}, \bar{x}_{ij}] \mid \bar{x}_{ij} \leq \bar{x}_{ij}, \bar{x}_{ij} \in \mathbb{R}\} \). Let \( y_i \) be the \( i \)th observation of the interval dependent variable \( Y \), where \( y_i = [\bar{y}_i, \bar{y}_i], \bar{y}_i \leq \bar{y}_i, y_i, y_i, y_i, y_i \in \mathbb{R} \).

2.1. CCRM Method. Let \( x_{ij}^{l} \) and \( x_{ij}^{r} \) be the midpoint and radius for the \( i \)th observation of the \( j \)th interval dependent variable, respectively, then equations (1) and (2) exist.

\[
x_{ij}^{l} = \frac{x_{ij} + \bar{x}_{ij}}{2},
\]

\[
x_{ij}^{r} = \frac{x_{ij} - \bar{x}_{ij}}{2}.
\]

Correspondingly, let \( y_{i}^{l} \) and \( y_{i}^{r} \) be the midpoint and radius for the \( i \)th observation of the \( j \)th interval dependent variable, respectively, then there exist equations (3) and (4) that hold.

\[
y_{i}^{l} = \frac{y_{i} + \bar{y}_{i}}{2},
\]

\[
y_{i}^{r} = \frac{\bar{y}_{i} - y_{i}}{2}.
\]
Based on formulas (1)–(4), an interval data series can be converted into two series, respectively, namely a midpoint and a radius. Let \( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p \) and \( \hat{\beta}_0', \hat{\beta}_1', \ldots, \hat{\beta}_p' \) be the regression coefficients of the midpoint and radius series, respectively, and \( \hat{\epsilon}_i \) as well as \( \hat{\epsilon}_i' \) be the corresponding errors. Then two linear regression formulas, namely, one for the midpoint series and another for the radius series holds. Mathematically, they can be written as equations (5) and (6), respectively.

\[
\begin{align*}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi} + \hat{\epsilon}_i, \\
\hat{y}_i' &= \hat{\beta}_0' + \hat{\beta}_1' x_{1i} + \cdots + \hat{\beta}_p' x_{pi}' + \hat{\epsilon}_i'.
\end{align*}
\]

According to the OLS method, the regression coefficients of above two regression models can be obtained by minimizing the sum of square of errors for midpoint and radius series. Mathematically, it can be expressed as equation (7).

\[
\min \sum_{i=1}^{n} \left[ (\hat{\epsilon}_i)^2 + (\hat{\epsilon}_i')^2 \right] = \sum_{i=1}^{n} \left[ (\hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_p x_{pi})^2 + (\hat{y}_i' - \hat{\beta}_0' + \hat{\beta}_1' x_{1i} + \cdots + \hat{\beta}_p' x_{pi}')^2 \right].
\]

Let \( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p \) and \( \hat{\beta}_0', \hat{\beta}_1', \ldots, \hat{\beta}_p' \) be the estimated regression coefficients. Then, the predicted values \( \hat{y}_i \) and \( \hat{y}_i' \) can be gained using equations (5) and (6). The interval data \( y_i = [\hat{y}_i, \hat{y}_i'] \) can be finally obtained by equation \( \hat{y}_i = \hat{y}_i - \hat{y}_i' \) and \( \hat{y}_i = \hat{y}_i + \hat{y}_i' \), respectively.

To apply the above model to the reality, the non-negativity of the values of the estimated radius must be added into the model because formula (7) could not assure the values of right endpoints greater than or equal to the values of the left endpoints [23]. Otherwise, as mentioned above, the magnitude of the left and right predicted intervals may be opposite in value, which is inconsistent with the reality.

2.2. **CCRM’ Method**. However, a certain number of evidences show that a considerable number of observed and predicted intervals become disjointed when the CCRM is applied. This phenomenon could be even worse when the fluctuation degree of the midpoint series is greater than that of the radius series. For this problem, Guo et al. [22] proposed a constraint center-and-range method in which two constraints were added into equation (8): (1) The right end value of the predicted interval is greater than or equal to the left end value of the corresponding observed interval, as shown in the first set of constraints (9); (2) The left end value of the predicted interval is less than or equal to the right end value of the corresponding observed interval, as shown in the second set of constraints (9). This improved CCRM model (called CCRM” for simplicity) can be expressed as (8), (9).

\[
\begin{align*}
\min \sum_{i=1}^{n} & \left[ (\hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_p x_{pi})^2 + (\hat{y}_i' - \hat{\beta}_0' + \hat{\beta}_1' x_{1i} + \cdots + \hat{\beta}_p' x_{pi}')^2 \right], \\
\text{s.t.} & \left\{ \begin{array}{l}
(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}) + (\hat{\beta}_0' + \hat{\beta}_1' x_{1i} + \cdots + \hat{\beta}_p' x_{pi}') \geq \hat{y}_i, \\
\hat{y}_i' + \hat{y}_i' \geq (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}) - (\hat{\beta}_0' + \hat{\beta}_1' x_{1i} + \cdots + \hat{\beta}_p' x_{pi}'), 
\end{array} \right. \\
& \forall i = 1, 2, \ldots, n.
\end{align*}
\]

3. **A Regression Method considering the Degree of Interval Coincidence**

To assure more intersections, the above CCRM+ may have to sacrifice the overall fitting accuracy in many situations (more detailed evidence can be seen in Monte Carlo simulation proposed by Guo [22]) and the result becomes even worse when the fluctuation degree of the midpoint data series is greater than that of the radius data series. It indicates that data singularities exist in an interval data series. When it happens, a certain number of disjointed elements may still exist by using the CCRM+, even though the overall fitting accuracy becomes worse. In this situation, the trade-offs between the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted interval and the overall fitting accuracy may have to be made by the decision-makers.

3.1. **Regression Model**. For an observation \( i \), in order to increase the coincidence degree between the observed and predicted intervals, the mathematical expression for such kind of a coincidence must be given in advance. Figure 1 shows three-couple relations between the observed and predicted intervals. The first couple (as shown in Figures 1(a) and 1(b)) illustrates a separation relation. The second and third couple show either intersection (as shown in Figures 1(c) and 1(e)) or containing relation (as shown in Figures 1(e) and 1(f)).
As can be seen, for an observation $i$, the maximization of such kind of a coincidence can be converted into the minimization of the distance between the observation and prediction of the left point, as well as the one of the right point. Mathematically, it can be expressed as formula (10).

\[
\min \left[ \left( y_i - \hat{y}_i \right)^2 + \left( \bar{y}_i - \tilde{y}_i \right)^2 \right], \quad \forall i \in \{1, 2, \ldots, n\}. \tag{10}
\]

where $\hat{y}_i$ and $\tilde{y}_i$ satisfy equation (11) and (12), respectively.

\[
\tilde{y}_i = y_i - \hat{y}_i = (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}), \tag{11}
\]

\[
\bar{y}_i = y_i + \hat{y}_i = (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) + (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}). \tag{12}
\]

Then, a novel nonlinear programming model can be presented, in which the maximization of the coincidence degree between observation $i$ and its prediction, as well as the minimization of the overall fitting error are both realized. Mathematically, the objective can be written as formula (13).

\[
\left\{ \rho \cdot \min_{1 \leq i \leq n} \left[ (y_i - \hat{y}_i)^2 + (\bar{y}_i - \tilde{y}_i)^2 \right] + (1 - \rho) \cdot \min_{i=1}^{n} \left[ (\epsilon_i)^2 + (\epsilon'_i)^2 \right] \right\}. \tag{13}
\]

where $\rho \in [0, 1]$ represents a regression preference coefficient and is used to adjust the decision preference of the maximization of the coincidence degree between an observation and its prediction, as well as the minimization of the overall fitting error.

\[
\left[ y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) + (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) \right] \geq \eta, \quad \forall i = 1, 2, \ldots, n. \tag{14}
\]

Based on it, the objective (13) can be rewritten as formula (15).

\[
\min \left\{ \rho \cdot \eta + (1 - \rho) \sum_{i=1}^{n} \left[ (y_i - \hat{y}_i)^2 + (y_i - \hat{y}_i)^2 \right] \right\}. \tag{15}
\]

In addition, constraints (16) must be added into the above model, ensuring that (1) the right end value of the predicted interval is greater than or equal to the left end value of the corresponding observed interval and (2) the left end value of the predicted interval is less than or equal to the right end value of the corresponding observed interval.

**Figure 1:** Schematic diagram of the relative relationship between observation and prediction interval.
shown in (19)–(22).

When $\rho = 0$, the first term in objective (15) is zero. The model degenerates to the CCRM$^+$ model. When $\rho = 1$, the second term in objective (15) is zero. The decision is converted into the maximization of the worst coincidence degree between all observations and their corresponding predictions.

3.2. Solving Regression Coefficients. For facilitating the explanation, the regression coefficients, error terms, as well as observation terms of the midpoint and radius of the interval data are re-expressed in matrix form (as shown below).

$\mathbf{\beta} = [\beta_0^c, \beta_1^c, \ldots, \beta_p^c]^T, \mathbf{\beta}' = [\beta_0^r, \beta_1^r, \ldots, \beta_p^r]^T, \beta = (\mathbf{\beta}^c, \mathbf{\beta}^r)^T,$

\[
\min \left\{ \rho \cdot \eta + (1 - \rho) \sum_{i=1}^{n} \left[ (\epsilon_i^c)^2 + (\epsilon_i^r)^2 \right] \right\},
\]

\[
\begin{aligned}
\text{s.t.} & \quad \left[ (y_i^c - \mathbf{X}^c \mathbf{\beta}^c + \mathbf{X}^r \mathbf{\beta}^r)^2 + (y_i^r - \mathbf{X}^c \mathbf{\beta}^c - \mathbf{X}^r \mathbf{\beta}^r)^2 \right] \geq \eta, \quad \forall i = 1, 2, \ldots, n, \\
\text{s.t.} & \quad y_i^c + y_i^r \geq \mathbf{X}^c \mathbf{\beta}^c - \mathbf{X}^r \mathbf{\beta}^r, \quad \forall i = 1, 2, \ldots, n, \\
\end{aligned}
\]

For the second term in objective (17), Guo et al. [22] had proved that it was a convex function. Associated with the linear characteristics of the first term in objective (17), the total objective function is a convex one. For each set of constraints in formula (18), we define several functions as shown in (19)–(22).

\[
g_{ii} = \left[ y_i^c - \mathbf{X}^c \mathbf{\beta}^c + \mathbf{X}^r \mathbf{\beta}^r \right]^2, \quad \forall i. \tag{19}
\]

\[
g_{ii} = \left[ y_i^r - \mathbf{X}^c \mathbf{\beta}^c - \mathbf{X}^r \mathbf{\beta}^r \right]^2, \quad \forall i. \tag{20}
\]

\[
g_{ii} = \mathbf{X}^c \mathbf{\beta}^c + \mathbf{X}^r \mathbf{\beta}^r - y_i^c + y_i^r, \quad \forall i. \tag{21}
\]

\[
\frac{g_{ii}}{g_{ii}} = y_i^c + y_i^r - \mathbf{X}^c \mathbf{\beta}^c + \mathbf{X}^r \mathbf{\beta}^r, \quad \forall i. \tag{22}
\]

Since equations (21) and (22) are linear functions, there exists the corresponding Hessian matrix $\mathbf{H}_{g_{ii}} = \mathbf{H}_{g_{ii}} = 0$ that holds. It means that both equations (21) and (22) are concave functions. For formula (19), we present the characteristic root determinant of the Hessian matrix as shown in formula (23).

\[
\frac{g_{ii}}{g_{ii}} = \left[ \begin{array}{cccc} 2 - \lambda & 2x_1^c & 2x_2^c & \cdots & 2x_p^c & -2 & -2x_1^c & -2x_2^c & \cdots & -2x_p^c \\
2x_1^c & (x_1^c)^2 - \lambda & 2x_1^c x_2^c & \cdots & 2x_1^c x_p^c & -2x_1^c & -2x_1^c x_2^c & \cdots & -2x_1^c x_p^c \\
2x_2^c & 2x_1^c x_2^c & (x_2^c)^2 - \lambda & \cdots & 2x_2^c x_p^c & -2x_2^c & -2x_2^c x_2^c & \cdots & -2x_2^c x_p^c \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
-2x_1^c & -2x_1^c x_2^c & -2x_2^c x_2^c & \cdots & (x_p^c)^2 - \lambda & -2x_1^c & -2x_1^c x_2^c & \cdots & -2x_1^c x_p^c \\
-2x_2^c & -2x_1^c x_2^c & 2x_1^c x_2^c & \cdots & (x_2^c)^2 - \lambda & -2x_2^c & -2x_2^c x_2^c & \cdots & -2x_2^c x_p^c \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
-2x_p^c & -2x_p^c x_2^c & -2x_2^c x_p^c & \cdots & (x_p^c)^2 - \lambda & \cdots & \cdots & \cdots & \cdots \\
\end{array} \right]_{(2p+2) \times (2p+2)}
\]

(23)
In order to obtain $|H_{g_{u1}}|$, determinant transformation has to be conducted. (1) For any row $i$ ($i = 1, 2, \ldots, p$), each element in the 1st row is multiplied by $-x_i^c$ ($i = 1, 2, \ldots, p$) and then be added to the corresponding element in row $i$. (2) Each element in the 1st row is added to the corresponding element in row $p + 1$. (3) For any row $i$ ($i = p + 2, p + 3, \ldots, 2p + 2$), each element in the 1st row is multiplied by $x_i^c$, ($i = 1, 2, \ldots, p$) and then be added to the corresponding element in row $i$. Then the transformed formula can be written as (24).

$$
|H_{g_{u1}}| = \begin{bmatrix}
2 - \lambda & 2x_1^c & 2x_2^c & \cdots & 2x_p^c & -2 & -2x_1^c & -2x_2^c & \cdots & -2x_p^c \\
\lambda x_1^c & -\lambda & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\lambda x_2^c & 0 & -\lambda & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\lambda x_p^c & 0 & 0 & \cdots & -\lambda & 0 & 0 & 0 & \cdots & 0 \\
-\lambda x_1^c & 0 & 0 & \cdots & 0 & 0 & -\lambda & 0 & \cdots & 0 \\
-\lambda x_2^c & 0 & 0 & \cdots & 0 & 0 & 0 & -\lambda & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-\lambda x_p^c & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -\lambda \\
\end{bmatrix}
$$

For any row $i$ ($i = 2, 3, \ldots, 2p + 2$), factor $\lambda$ can be extracted to form equation (25).

$$
|H_{g_{u1}}| = \lambda^{(2p+1)}
$$

Furthermore, (1) for any row $i$ ($i = 2, 3, \ldots, p$), each element in the $i$th row multiples $2x_i^c$ and then be added to the corresponding element in row 1. (2) Each element in the $(p+1)$th row multiples $(-2)$ and then be added to the corresponding element in row 1. (3) For any row $i$ ($i = p + 2, p + 3, \ldots, 2p + 2$), each element in the $i$th row multiples $(-2x_i^c)$ and then be added to the corresponding row 1. The transformed formula can be further rewritten as equation (26).

Formula (26) is a lower triangular matrix, so the determinant with characteristic roots can be transformed as equation (27). Finally, the characteristic root determinant of the Hessian matrix is $\lambda = 0$ or $\lambda = 4 + \left(\sum_{i=1}^{p} 2 (x_i^c)^2 + \sum_{i=1}^{p} + 12 (x_i^c)^2\right) \geq 0$. It means that $H_{g_{u1}}$ is a positive semi-definite.
matrix and \( g_1 \) is a concave function. Similarly, \( g_2 \) is also a concave function, because a linear combination of two concave functions is still a concave function. Therefore, models (17) and (18) constitute a convex programming model, which can be solved by K–T condition.

\[
\begin{array}{c}
\left| H_{g_{\lambda}} \right| = \lambda^{(2p+1)} \begin{bmatrix}
4 + \left( \sum_{i=1}^{p} 2(x_i^2) + \sum_{i=1}^{p+1} 2(x_i^2) \right) - \lambda & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\
-1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & -1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & -1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & -1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & -1 & 0 & \ldots & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & -1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 0 & 0 & -1 & \ldots & 0 \\
\end{bmatrix} , (26)
\end{array}
\]

\[
\left| H_{g_{\lambda}} \right| = (-1)^{2p+1} \lambda^{(2p+1)} \begin{bmatrix}
4 + \left( \sum_{i=1}^{p} 2(x_i^2) + \sum_{i=1}^{p+1} 2(x_i^2) \right) - \lambda \\
\end{bmatrix} = 0. \tag{27}
\]

3.3. Evaluation Function. To evaluate the overall fitting accuracy, this paper uses RSMR [25, 26] to express mathematically as formula (28).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} DM(y_i, \bar{y}_i)^2}{n}} \tag{28}
\]

where \( DM(y_i, \bar{y}_i) = |y_i^c - \bar{y}_i^c| + |y_i^e - \bar{y}_i^e| \), denoting Hausdorff distance. \( y_i^c \) and \( y_i^e \) represent an observation and its prediction value of center’s dependent variable, respectively. \( \bar{y}_i^c \) and \( \bar{y}_i^e \) represent an observation and its prediction value of radius’ dependent variable, respectively. \( n \) is the number of observations.

To evaluate the coincidence degree of the observed and their predicted intervals, three indicators are used [27].

(1) AR. It consists of a series of added fractions on average, each of which is an intersection width between an observed and its predicted interval to the union width between the observed and its predicted interval. Mathematically, it can be expressed as formula (29).

\[
AR = \frac{1}{n} \sum_{i=1}^{n} \frac{w(y_i \cap \bar{y}_i)}{w(y_i \cup \bar{y}_i)} \tag{29}
\]

where \( w(\cdot) \) represents the width of an interval.

(2) PCO. It consists of a series of added fractions on average, each of which is an intersection width between an observed and its predicted interval to the width of the observed interval. Mathematically, it can be expressed as formula (30).

\[
PCO = \frac{1}{n} \sum_{i=1}^{n} \frac{w(y_i \cap \bar{y}_i)}{w(y_i)}. \tag{30}
\]

(3) N0. The average number of forecasts with 0% accuracy. It means the average number of disjointed elements between all observed and their predicted intervals among all used testing sets of interval series.

4. Monte Carlo Simulation Analysis

4.1. Generation of Simulation Data. In order to illustrate the advantage of the proposed model, comparisons with the above CCRM and CCRM+ are conducted as follows:

(1) To illustrate the influence of the number of interval independent variables on the fitting effect, two cases are simulated, in which the number of independent variables are \( j = 1 \) and \( j = 4 \), respectively.

(2) To express the relation between the midpoint of the interval data and the fluctuation degree of the radius of the interval data, different types of error terms are used to form three regression models.

(3) To illustrate the impact of preference choice on regression effect, two scenarios, namely, \( \rho = 0 \) and \( \rho = 1 \) are considered into this simulation.

Table 1 shows 6 scenarios in details. The generation process for simulated data could be described as follows:
Based on the assumption that the decision-maker requires the maximum coincidence degree of samples with the worst coincidence degree between the observed and predicted intervals, the hypothesis test results of the comparisons between the proposed model and the CCRM+ are shown in Tables 2-4.

The above conclusions can be pointed out from C1–C3.

(1) The root mean square error rejects the 0 hypothesis that "the proposed model is not inferior to the CCRM+" and "the proposed model is equivalent to the CCRM+" but accepts "the CCRM+ is not inferior to the proposed model". This suggests that the proposed model is inferior to the CCRM+ in the overall fitting accuracy.

(2) Under the zero assumption that "the proposed model is not inferior to the CCRM+" and "the proposed model is equivalent to the CCRM+", the rejection rate of the average accuracy rate (AR) is about 50%-60%. This suggests that the proposed model is equivalent to the CCRM+ in terms of average accuracy.

(3) The average ratio (PCO) of the observation interval including the prediction interval rejects the 0 hypothesis that "the CCRM+ is not inferior to the proposed model" and "the proposed model is equivalent to the CCRM+" but accepts the 0 hypothesis that "the proposed model is not inferior to the CCRM+". This suggests that the proposed model is more advantageous in that the observation interval contains the average ratio of the prediction interval.

When the objective of a decision-maker is to maximize the coincidence degree of the sample with the worst coincidence degree between the observed and predicted intervals, the proposed model has a larger average ratio of all predicted intervals contained within their observed intervals, but this is at the expense of the overall fitting accuracy. Table 5 shows the average value of various evaluation indices under 100 sets of test data. Compared with the CCRM+, the proposed model improves the accuracy of the PCO by sacrificing a smaller RMSE, especially when there are more independent variables as well as the degree of midpoint fluctuation is larger than the degree of radius fluctuation (as shown in the 8th column of Table 5), the advantage of the proposed model is more obvious in terms of PCO.

Table 6 shows the statistical data related to the number of samples whose prediction accuracy is 0.
Figure 2: Randomly generating data histogram in $C_1$.

Figure 3: Randomly generating data histogram in $C_2$.

Figure 4: Randomly generating data histogram in $C_3$. 

Compared to the CCRM, the proposed model and the CCRM+ can obtain a lower number of samples with a prediction accuracy of 0. When the independent variable \( j = 1 \), the number of samples with 0 prediction accuracy resulting from the proposed model under \( C_1-C_3 \) are all zero. \( \sqrt{X_h} \) suggests that the proposed model can make all of the observed intervals and predicted intervals intersect. When the independent variable \( j = 4 \), the use of the proposed model involves a number of intersections between the observed and predicted intervals, but it is still better than the corresponding number obtained by the CCRM+. Note that when the fluctuation degree of the midpoint data series is higher than the radius fluctuation degree, such an effect from the proposed model is more significant.

Based on the assumption that the decision-maker requires the minimization of the sum of squares of the overall error for both midpoint and radius series, the hypothesis test results on the comparison of the proposed model to the CCRM are as shown in Tables 7–9.

The above conclusions can be pointed out from \( C_4-C_6 \).

(1) The root mean square error rejects the 0 hypothesis that “the proposed model is not inferior to the CCRM” and “the proposed model is equivalent to the CCRM” but accepts “the CCRM is not inferior to the proposed model”. This suggests that the proposed model is inferior to the CCRM in the overall fitting accuracy.

(2) The average accuracy rate (AR) rejects the 0 hypothesis that “the proposed model is not inferior to the CCRM” and “the proposed model is equivalent to the CCRM” but accepts “the CCRM is not inferior to the proposed model”. This suggests that the

| Table 2: Experimental results on null hypothesis of the proposed model are not inferior to CCRM+. |
|---|---|---|---|
| Number | \( j \) | RMSE (%) | AR (%) |
| \( C_1 \) | \( j = 1 \) | 90 | 55 |
| | \( j = 4 \) | 92 | 56 |
| \( C_2 \) | \( j = 1 \) | 94 | 58 |
| | \( j = 4 \) | 85 | 48 |
| \( C_3 \) | \( j = 1 \) | 91 | 54 |
| | \( j = 4 \) | 90 | 56 |

| Table 3: Experimental results on null hypothesis of CCRM+ are not inferior to the proposed model. |
|---|---|---|---|
| Number | \( j \) | RMSE (%) | AR (%) |
| \( C_1 \) | \( j = 1 \) | 4 | 7 |
| | \( j = 4 \) | 7 | 3 |
| \( C_2 \) | \( j = 1 \) | 2 | 3 |
| | \( j = 4 \) | 8 | 0 |
| \( C_3 \) | \( j = 1 \) | 8 | 2 |
| | \( j = 4 \) | 7 | 5 |

| Table 4: Experimental results on null hypothesis of the proposed model are equivalent to CCRM+. |
|---|---|---|---|
| Number | \( j \) | RMSE (%) | AR (%) |
| \( C_1 \) | \( j = 1 \) | 100 | 48 |
| | \( j = 4 \) | 100 | 52 |
| \( C_2 \) | \( j = 1 \) | 100 | 51 |
| | \( j = 4 \) | 100 | 55 |
| \( C_3 \) | \( j = 1 \) | 100 | 52 |
| | \( j = 4 \) | 100 | 56 |

| Table 5: Calculating results of evaluation indices on testing set. |
|---|---|---|---|---|---|---|---|
| Number | Model | \( j = 1 \) | \( j = 4 \) |
| | | RMSE | AR | PCO | RMSE | AR | PCO |
| \( C_1 \) | CCRM+ | 121.211 | 0.340 | 0.917 | 521.541 | 0.210 | 0.857 |
| Proposed model | 123.126 | 0.331 | 0.968 | 533.050 | 0.208 | 0.904 |
| \( C_2 \) | CCRM+ | 128.688 | 0.335 | 0.901 | 527.874 | 0.203 | 0.875 |
| Proposed model | 132.648 | 0.330 | 0.953 | 540.490 | 0.188 | 0.928 |
| \( C_3 \) | CCRM+ | 167.94 | 0.292 | 0.917 | 542.214 | 0.182 | 0.905 |
| Proposed model | 182.358 | 0.273 | 0.945 | 554.327 | 0.173 | 0.949 |
proposed model is inferior to the CCRM in terms of average accuracy.

(3) The average ratio (PCO) of the observed interval including the predicted interval rejects the 0 hypothesis that "the CCRM is not inferior to the proposed model" and "the proposed model is equivalent to the CCRM" but accepts the 0 hypothesis that "the proposed model is not inferior to the CCRM". This suggests that the proposed model is more advantageous in that the observation interval contains the average ratio of the prediction interval.

When the objective of a decision-maker is to minimize the sum of squares of the overall error for both midpoint and radius series, the proposed model has a better average ratio of the observation interval including the prediction interval, but at the expense of the overall fitting accuracy. Table 10 shows the average value of the evaluation index under 100 sets of test data. As can be seen, compared with the CCRM, the proposed model has significantly improved the accuracy of the PCO by sacrificing smaller RMSE and AR values; especially when the fluctuation degree of the midpoint data series is higher than the radius fluctuation degree, the effect of the proposed model is more significant.

Table 11 shows the statistical data related to the number of samples with a prediction accuracy of 0. As can be seen, as the fluctuation degree of the midpoint series increases, the number of samples with a prediction accuracy of 0 for the CCRM gradually increases, but the proposed model maintains a relatively low level. This suggests that the proposed model has a significant advantage in the intersection of the observed and predicted intervals formed in the fitting process.

5. Empirical Example

To illustrate the application, the relation between civil aviation passenger turn- over and GDP in China is studied. The related data come from the monthly air passenger turnover data and GDP quarterly data released by the National Bureau of Statistics in China (http://www.stats.gov.cn/tjss/,2021) [28]. The quarterly air passenger turnover and GDP from 2005 to 2019 are selected as the sample. The minimum value of four quarterly data in each year is used as the lower limit of the interval data, and the maximum value is used as the upper limit of the interval number to construct interval data, respectively. The statistics data obtained are as shown in Table 12.

Figure 5 shows the trend of this interval series, in which it appears to be a linear relation and the size of these intervals becomes bigger and bigger.

As can be seen from Table 13, compared to the CCRM, although the proposed model degenerates into a kind of a CCRM, this proposed model has an improvement on the interval coincidence degree by slightly sacrificing the overall fitting accuracy. Compared to the CCRM+, due to the proposed model degenerating to the CCRM+, the results show an equivalent effect in terms of RMSE, AR, PCO, and N0, which is consistent with the Monte Carlo simulation.

With the increase of the regression preferencep, the first term in objective (15), namely, the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted intervals, becomes gradually important. In this situation, a certain number of predicted intervals may have to enlarge their widths to achieve this goal. This would lead to a raise of RMSE (as shown in Figure 6). This kind of an enlargement not only causes the increase of unions between the observed and predicted intervals but also leads to the decrease of AR (as shown in Figure 7). Because of the forced intersection, the intersection between the observed and predicted intervals is also enlarged, which leads to the rise of PCO (as shown in Figure 8).

Note that all of the changes in Figures 6–8 are not with sustainable growth and occur with some slight fluctuations. Taking $\rho = 0$ and $\rho = 0.1$ as examples, the RMSE decreases but again increases at $\rho = 0.2$. The reason lies in the fact that with the increase of $\rho$, the maximization of coincidence degree for the sample with the worst coincidence degree between the observed and predicted intervals gradually becomes important in objective (15).

In order to ensure the intersection of the worst sample, the decision variables are changed wherein when the samples with the total reducing values of RMSE are greater than the ones of the total increasing values of RMSE, the final result decreases and vice versa (as shown in Figure 6).

To illustrate the advantage of the proposed model when some data singularities exist, we add the interval data of 2020s (as the dotted and bordered rectangle shown in Figure 9), wherein the passenger turnover as well as the GDP have sharply dropped owing to the outbreak of the COVID-19 pandemic (as shown in Table 12).

Considering the impact of the regression preference coefficient as well as its effect as shown above, we set $\rho = 0.9$ (determined by tests) to conduct this comparison (as shown in Table 14).

The compared objectives are still the CCRM and CCRM+. The results show that the use of the proposed model can generate a smaller number of N0 than the use of the CCRM. In addition, compared to the use of CCRM+, the proposed model significantly increases the value of PCO, whereas with a slight decrease of RMSE and AR, as explained...
### Table 7: Experimental results on null hypothesis of the proposed model are not inferior to CCRM.

| Number | Number | RMSE (%) | AR (%) | PCO (%) |
|--------|--------|----------|--------|---------|
| C_4    | j=1    | 100      | 98     | 0       |
|        | j=4    | 100      | 92     | 0       |
| C_5    | j=1    | 100      | 90     | 0       |
|        | j=4    | 100      | 95     | 0       |
| C_6    | j=1    | 100      | 95     | 0       |
|        | j=4    | 100      | 91     | 0       |

### Table 8: Experimental results on null hypothesis of CCRM are not inferior to the proposed model.

| Number | Number | RMSE (%) | AR (%) | PCO (%) |
|--------|--------|----------|--------|---------|
| C_4    | j=1    | 0        | 3      | 99      |
|        | j=4    | 0        | 4      | 98      |
| C_5    | j=1    | 0        | 2      | 100     |
|        | j=4    | 0        | 2      | 97      |
| C_6    | j=1    | 0        | 3      | 99      |
|        | j=4    | 0        | 2      | 98      |

### Table 9: Experimental results on null hypothesis of the proposed model are equivalent to CCRM.

| Number | Number | RMSE (%) | AR (%) | PCO (%) |
|--------|--------|----------|--------|---------|
| C_4    | j=1    | 100      | 97     | 100     |
|        | j=4    | 100      | 98     | 100     |
| C_5    | j=1    | 100      | 100    | 100     |
|        | j=4    | 100      | 97     | 100     |
| C_6    | j=1    | 100      | 100    | 100     |
|        | j=4    | 100      | 97     | 100     |

### Table 10: Calculating results of evaluation indices on the testing set.

| Number | Model      | RMSE | AR  | PCO  | RMSE | AR  | PCO  |
|--------|------------|------|-----|------|------|-----|------|
| C_4    | CCRM       | 81.295 | 0.377 | 0.495 | 494.364 | 0.242 | 0.375 |
|        | Proposed model | 110.37 | 0.352 | 0.905 | 505.967 | 0.228 | 0.904 |
| C_5    | CCRM       | 109.014 | 0.352 | 0.429 | 517.520 | 0.212 | 0.285 |
|        | Proposed model | 139.295 | 0.337 | 0.892 | 510.825 | 0.196 | 0.918 |
| C_6    | CCRM       | 136.654 | 0.347 | 0.405 | 539.260 | 0.193 | 0.244 |
|        | Proposed model | 193.691 | 0.316 | 0.912 | 552.524 | 0.182 | 0.929 |

### Table 11: Statistical results on sample counts (N0) of zero-prediction accuracy rate.

| Preference coefficient (j = 1) | CCRM | Proposed model | Preference coefficient (j = 4) | CCRM | Proposed model |
|-------------------------------|------|----------------|-------------------------------|------|----------------|
| C_4                           | 20   | 0              | C_4                           | 27   | 1              |
| C_5                           | 28   | 0              | C_5                           | 35   | 1              |
| C_6                           | 38   | 1              | C_6                           | 40   | 2              |
Table 12: The maximum and minimum data of civil aviation passenger turnover and GDP quarterly.

| Year | Minimum quarterly GDP (trillion yuan) | Maximum quarterly GDP (trillion yuan) | Minimum quarterly turnover volume (100 million person-kilometers) | Maximum quarterly turnover volume (100 million person-kilometers) |
|------|--------------------------------------|---------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| 2005 | 40453.30                             | 54024.80                              | 432.97                                                        | 573.75                                                        |
| 2006 | 47078.90                             | 63621.60                              | 521.12                                                        | 669.02                                                        |
| 2007 | 57159.30                             | 78669.30                              | 616.92                                                        | 775.37                                                        |
| 2008 | 69373.60                             | 88699.00                              | 685.22                                                        | 751.65                                                        |
| 2009 | 73979.20                             | 100825.80                             | 783.69                                                        | 906.87                                                        |
| 2010 | 87501.30                             | 119306.80                             | 944.99                                                        | 1111.54                                                       |
| 2011 | 104469.90                            | 138012.10                             | 1048.68                                                       | 1218.56                                                       |
| 2012 | 117357.60                            | 151812.00                             | 1180.22                                                       | 1375.13                                                       |
| 2013 | 129449.60                            | 167772.30                             | 1331.82                                                       | 1558.79                                                       |
| 2014 | 140759.80                            | 180828.90                             | 1488.10                                                       | 1708.30                                                       |
| 2015 | 151137.90                            | 192572.90                             | 1722.62                                                       | 1970.69                                                       |
| 2016 | 162410.00                            | 211566.20                             | 1979.42                                                       | 2259.23                                                       |
| 2017 | 181867.70                            | 235428.70                             | 2274.92                                                       | 2511.28                                                       |
| 2018 | 202035.70                            | 258808.90                             | 2574.07                                                       | 2813.42                                                       |
| 2019 | 217168.30                            | 276798.00                             | 2853.49                                                       | 3114.24                                                       |
| 2020 | 205727.00                            | 296297.08                             | 1058.06                                                       | 2012.28                                                       |

Figure 5: Rectangular graph of interval samples of the passenger turnover and GDP.

Table 13: Results of evaluation indexes of regression prediction effect of different methods.

| Indicator | RMSE  | AR    | PCO   | N0 |
|-----------|-------|-------|-------|----|
| Method proposed model (0, 1)    | 119.882 | 0.560 | 0.698 | 0  |
| CCRM indicator                  | 117.457 | 0.580 | 0.679 | 2  |
| Method proposed model (1, 0)    | 134.923 | 0.440 | 0.887 | N0 |
| CCRM+                              | 119.882 | 0.560 | 0.698 | 0  |

Note that "∗" in Proposed model (∗,1 − ∗) represents the value of $\rho$ in the proposed model.
Figure 6: RMSE of the three models.

Figure 7: AR of the three models.

Figure 8: PCO of the three models.
above, to some extent, by enlarging the widths of some predicted intervals. It suggests that using the proposed model can obtain a better solution.

6. Conclusion

By adding a nonnegativity condition to the predicted radius, traditional CCRM solves the consistency problem of the prediction result of the right and left endpoints. But CCRM could not guarantee that all observed and predicted intervals will intersect. Therefore, two constraints, considering that (1) the right end value of the predicted interval is greater than or equal to the left end value of the corresponding observed interval, as well as (2) the left end value of the predicted interval is less than or equal to the right end value of the corresponding observed interval, are added into the CCRM model. This so-called CCRM+ model outperforms the use of the CCRM model in terms of N0. However, this model still has a poor performance when several singularities exist in interval data series. In fact, the CCRM+ model is to increase the coincidence degree at the expense of decreasing the overall fitting accuracy. Namely, the model could not make trade-offs between them, which to some extent is of great importance for decision-makers to adjust the result to their expectation.

This paper presents a novel programming model, in which the minimization of the sum of square of errors for both center and range sample series, as well as the maximization of coincidence degree for the sample with the worst coincidence degree between observed and predicted intervals are taken into account and can be flexibly adjusted using a regression preference coefficient. This proposed model is particularly useful when the fluctuation degree of the midpoint data series is greater than that of the radius data series. For solution algorithm, the presented model is proven to be a concave one that satisfied K–T condition. Monte Carlo simulation shows that the proposed model has a better performance on the average percentage of predicted intervals contained in the observed intervals, as well as the average number of forecasts with 0% accuracy. For application, airline industry data are used to illustrate the value of the proposed model. Note that some coincidences within observations of interval data are realized by enlarging the radius of the prediction intervals. This actually increases the fitting error. Thus, determining how to increase the coincidence degree of the observed and predicted intervals without increasing the predicted intervals’ radius is the next important research direction.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work that is described in this paper has been supported by the National Science Foundation of China (Grant
References

[1] T. Xiong, C. Li, Y. Bao, Z. Hu, and L. Zhang, "A combination method for interval forecasting of agricultural commodity futures prices," Knowledge-Based Systems, vol. 77, no. 3, pp. 92–102, 2015.

[2] M. Amagasa and K. Nagata, "Prediction model with interval data -toward practical applications," in Proceedings of the IPMU 2016: Information Processing and Management of Uncertainty in Knowledge-Based Systems, vol. 611, pp. 213–224, Springer, Eindhoven, The Netherlands, June 2016.

[3] X. Wang, S. Li, and T. Denneux, "Interval-valued linear model," International Journal of Computational Intelligence Systems, vol. 8, no. 1, pp. 114–127, 2015.

[4] I. Izonin, R. Tkachenko, N. Kryvinska, P. Tkachenko, and M. Greguis, "Multiple linear regression based on coefficients identification using non-iterative SGTM neural-like structure," in Proceedings of the International Work-Conference on Artificial Neural Networks, pp. 467–479, Springer, Gran Canaria, Spain, June 2019.

[5] J. Jang and K.-H. Kang, "Local linear regression analysis for interval-valued data," Communications for Statistical Applications and Methods, vol. 27, no. 3, pp. 365–376, 2020.

[6] R. Boukezzoulia, S. Galichet, and A. Bissierier, "A Midpoint-Radius approach to regression with interval data," International Journal of Approximate Reasoning, vol. 52, no. 9, pp. 1257–1271, 2011.

[7] L. Billard and E. Diday, "Symbolic regression analysis," in Proceedings of the Classification, Clustering and Data Analysis, Proceedings of the 8th Conference of the International Federation of Classification Societies, pp. 281–288, Springer, Poland, January 2002.

[8] L. C. Souza, R. M. C. R. Souza, G. J. A. Amaral, and T. M. Silva Filho, "A parametrized approach for linear regression of interval data," Knowledge-Based Systems, vol. 131, pp. 149–159, 2017.

[9] L. Billard and E. Diday, "Regression analysis for interval-valued data," in Proceedings of Data Analysis, Classification, and Related Methods-Proceedings (IFCS2000), pp. 369–374, Springer, Heidelberg, Germany, June 2000.

[10] E. d. A. Lima Neto and F. d. A. T. de Carvalho, "Centre and Range method for fitting a linear regression model to symbolic interval data," Computational Statistics & Data Analysis, vol. 52, no. 3, pp. 1500–1515, 2008.

[11] C. Wei, N. Zheng, and K. Tian, "Efficient estimation of interval-valued symbolic data regression model," International Journal of Statistics and Probability, vol. 9, no. 23, pp. 1–6, 2020.

[12] Y Zhong, Z. Zhang, and S. Li, "A constrained interval-valued linear regression model: a new heteroscedasticity estimation method," Journal of Systems Science and Complexity, vol. 33, no. 6, pp. 333–351, 2020.

[13] J. E. Chacón and O. Rodríguez, "Regression models for symbolic interval-valued variables," Entropy, vol. 23, no. 4, p. 429, 2021.

[14] C. Lim, "Interval-valued data regression using nonparametric additive models," Journal of the Korean Surgical Society, vol. 45, no. 3, pp. 358–370, 2016.

[15] E. A. Lima Neto and F. A. T. de Carvalho, "Constrained linear regression models for symbolic interval-valued variables," Computational Statistics & Data Analysis, vol. 52, no. 4, pp. 333–347, 2010.

[16] W. Huwén, R. Guan, and W. Junjie, "Linear regression of interval-valued data based on complete information in hypercubes," Systems Engineering Society of China, vol. 21, no. 4, pp. 422–442, 2012.

[17] J. Guo and P. Hao, "Constrained center and range joint model for interval-valued symbolic data regression," Computational Statistics & Data Analysis, vol. 116, 2017.

[18] Z. G. Su, P. H. Wang, Y. G. Li, and Z. K. Zhou, "Parameter estimation from interval-valued data using the expectation-maximization algorithm," Journal of Statistical Computation and Simulation, vol. 85, no. 1-3, pp. 320–338, 2015.

[19] W. Xun, Z. Zhang, and S. Li, "Set-valued and interval-valued stationary time series," Journal of Multivariate Analysis, vol. 145, pp. 208–223, 2016.

[20] R. Boukezzoula, S. Galichet, and D. Coquin, "From fuzzy regression to gradual regression: interval-based analysis and extensions," Information Sciences, vol. 441, pp. 18–40, 2018.

[21] E. d. A. L. Neto and F. D. A. T. De Carvalho, "An exponential-type kernel robust regression model for interval-valued variables," Information Sciences, vol. 454-455, pp. 419–442, 2018.

[22] J. P. Guo, R. Zhao, and W. H. Li, "A constrained CRM interval regression method," Journal of Management in Engineering, vol. 30, no. 4, pp. 196–202, 2016.

[23] E. D. A. Lima Neto and F. D. A. T. De Carvalho, "Nonlinear regression applied to interval-valued data," Pattern Analysis & Applications, vol. 20, no. 3, pp. 1–16, 2017.

[24] P. Giordani, "Lasso-constrained regression analysis for interval-valued data," Advances in Data Analysis and Classification, vol. 9, no. 1, pp. 5–19, 2015.

[25] C.-C. Chuang, "Extended support vector interval regression networks for interval input-output data," Information Sciences, vol. 178, no. 3, pp. 871–891, 2008.

[26] F. d. A. T. De Carvalho, R. M. C. R. De Souza, M. Chavent, and Y. Lechevallier, "Adaptive Hausdorff distances and dynamic clustering of symbolic interval data," Pattern Recognition Letters, vol. 27, no. 3, pp. 167–179, 2006.

[27] W. H. Li and J. P. Guo, "Interval symbolic data regression analysis and its application," Journal of Management Science, vol. 13, no. 4, pp. 38–44, 2010.

[28] National Bureau of Statistics, Monthly Data, http://www.stats.gov.cn/tjsj/, 2021.