A Method for Predicting the Macro Environment of Enterprise Management Based on Linear Models

Zhengxuan Hu *

School of Jiangxi Agricultural University, Jiangxi, China

* Corresponding author: 1224786480@qq.com

Abstract. Aiming at interval data, this paper combines the additive model on the basis of the partial linear model and introduces the sliding window model, and proposes an additive linear model based on the midpoint and range of each window of the sliding window mode. This model combines the advantages of semi-parametric regression and sliding window model, and avoids the dimensional disaster. In addition, the model determines the number of phases of the sliding window according to the cross-entropy criterion, and gives an iterative algorithm for estimating model parameters and unknown function based on the least square method and kernel estimation method. In the empirical analysis, several financial indicators are introduced to predict the law of macroeconomic development. The results show that the improved model is better than the traditional regression model.

Keywords: enterprise management, macro environment, linear model, predicative methods, big data.

1. Introduction

Interval data is widely used in economic, financial and social life. It describes the variation range of variables and has the advantage of more abundant data information. Moreover, due to the uncertainties in real problems, it is difficult to obtain accurate point value prediction, and even if it can be obtained, the information contained in point value data is very limited. Therefore, the prediction results based on interval data have more reference value, and have been widely concerned by scholars at home and abroad.

Moore (1966) [1] solved the problem of uncertain values of variables in mathematical models through interval analysis, making the model results more reliable. Although there are many kinds of modeling methods for interval data analysis, most of them study the model method under the attribute of interval data point value by performing interval operation on interval data. Billard and Diday (2000) [2] first proposed the method of data midpoint (CM) for fitting regression of interval data. Billard and Diday (2003) [3] also proposed the max-min method, which converted the interval type data into the point value data of the boundary, and then carried out fitting prediction on the point value data of the interval boundary. Neto and Carvalho (2004, 2008) [4, 5] proposed the midpoint and range method (CRM), which transformed the original interval data into two-point value variables, center and range, and built models for the center and range of the interval respectively. Meanwhile, the midpoint and range method, midpoint method and max-min method are compared based on Monte Carlo simulation, and the results show that the prediction effect of midpoint and range method is better than that of midpoint method and max-min method. Meanwhile, the midpoint and range method, midpoint method and max-min method are compared based on Monte Carlo simulation, and the results show that the prediction effect of midpoint and range method is better than that of midpoint method and max-min method. Yang Wei (2016) [6] built a range high price and range low price model based on the high and low-price data of the financial time series, and conducted an empirical analysis of the financial range time series of the Chinese and American stock markets. Lim (2016) [7] proposed a midpoint range non-parametric additive model (CRAM), which is characterized by its shape, the formula is as follows:

$$Y_i^c = u^c + \sum_{j=1}^{p} f_j^c (X_{ij}^c) + \quad (1)$$

$$Y_i^r = u^r + \sum_{j=1}^{p} f_j^r (X_{ij}^r) + \varepsilon_i^r \quad (2)$$

Where, u^c and u^r are the intercept terms; f_j^c and f_j^r are the unknown smooth functions; ε_i^c and ε_i^r are independent equally distributed random errors with a mean of 0 and a finite variance, σ_c^2 and σ_r^2 respectively. f_j^c and f_j^r can be estimated by the backfitting algorithm proposed by Buja (1989) [8] and the empirical results show that the non-parametric additive model has a better fitting effect than the linear model.

The sliding window model has the advantage of selecting data with less volatility for predictive analysis under different window periods. In the study of sliding window model, Tao Zhifu (2020) [9] studied a class of non-negative variable weight combination prediction methods from the perspective of sliding window, and the example analysis shows that the introduction of sliding window technology can effectively improve the combination prediction accuracy. Combining semi-parametric regression, additive model and sliding window model, a partially additive linear model of midpoint range based on sliding window is proposed in this paper.

2. Fundamentals of mathematical model

2.1. Interval number related concepts and algorithms

Definition 1[10]: The measurable mapping $Y: \omega \to I_R$ on the probability space (ω, F, P) is defined as the extended random interval Y , where, $I_R$is the space formed by the ordered set of numbers on R. Simplistically, the extended random interval is $\widetilde{Y}(\omega) = [Y_L(\omega), Y_R(\omega)]$, where any $\omega \in R, Y_L(\omega) \in R$ and $Y_R(\omega) \in R$ are called the left boundary and right boundary, respectively.

If$\widetilde{A} = [A_L, A_R], \widetilde{B} = [B_L, B_R]$ is any two random intervals,$\lambda \in R, \lambda \geq 0$ is any scalar, then the algebraic operation of the random interval is defined as follows:

(1) The addition operation: $\widetilde{A} + \widetilde{B} = [A_L + B_L, A_R + B_R]$.

(2) Number multiplication operation:$\lambda \cdot \widetilde{A} = [\lambda \cdot A_L, \lambda \cdot A_R]$.

(3) Difference operation [11]: $\widetilde{A} -_H \widetilde{B} = [A_L - B_L, A_R - B_R]$.

In addition to defining the upper limit and lower limit representation of 1, the random interval can also be in the form of the center and radius of the interval, i.e.,$\widetilde{Y}(\omega) = [c(\omega), r(\omega)]$, were,$c(\omega) = \frac{(Y_L(\omega)+Y_R(\omega))}{2}, r(\omega) = \frac{(Y_R(\omega)-Y_L(\omega))}{2}$.

Definition 2[12]: if$\{Y_t = [Y_{L,t}, Y_{R,t}] = (c_t, r_t), t = 1,2,\ldots,n\}$,then $\{Y_t\}$ is called an interval time series.

2.2. Sliding window model

The sliding window model based on the fixed window length is to frame the observation sequence according to the specified unit length, and the sliding window is constantly updated in the unit of the fixed window, so as to calculate the statistical indicators in the box. By combining the observation time series with the sliding window model, the local observation time series can be independently observed and processed, and the time series can be analyzed on a microscopic scale [9]. Figure 1 shows a sliding window model of a time series with a window length of m.

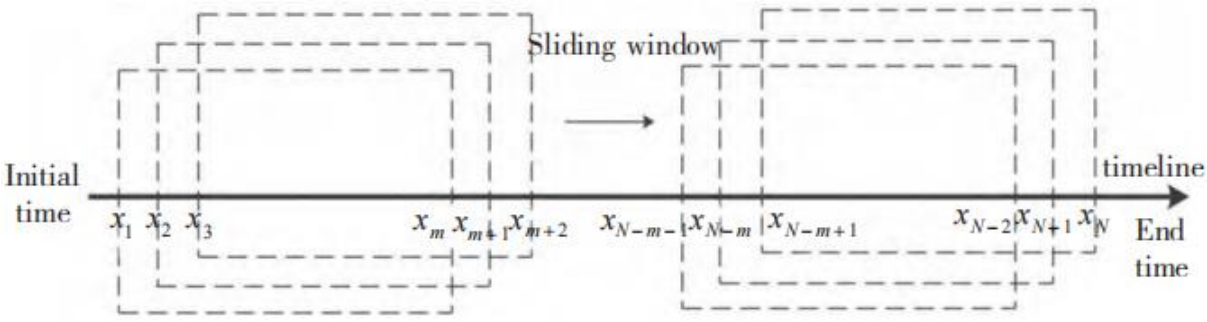

Figure 1. a time series sliding window model with window length m

3. A partially additive linear model of midpoint range based on sliding window

3.1. Determination of window period

Under different window periods, the fluctuation characteristics of data distribution are different, so it will be more meaningful to select the data with less volatility for prediction analysis. In this paper, the sliding window model is combined with the partially additive linear model of interval data to provide a new analytical perspective for the prediction of interval data. Firstly, the window period of the sliding window model is determined, and the following structure is considered:

Step 1: Set the number of window periods as m, and normalize the original center sequence y_t^c and the original range sequence y_t^r at different times:

$$p_t^c = y_t^c / \sum_{t=1}^{N-m+1} y_t^c, t = 1,2,\ldots,N-m+1 \tag{3}$$

$$p_t^r = y_t^r / \sum_{t=1}^{N-m+1} y_t^r, t = 1,2,\ldots,N-m+1 \tag{4}$$

Step 2: Calculate the original center and range under different number of window periods respectively cross entropy of the sequence:

$$h_m^c = \sum_{t_1=1}^{N-m+1} \sum_{t_2=1}^{N-m+1} p_{t_1}^c \ln\left(\frac{p_{t_1}^c}{p_{t_2}^c}\right) \tag{5}$$

$$h_m^r = \sum_{t_1=1}^{N-m+1} \sum_{t_2=1}^{N-m+1} p_{t_1}^r \ln\left(\frac{p_{t_1}^r}{p_{t_2}^r}\right) \tag{6}$$

Step 3: Select the window period m with the minimum cross entropy of the center and range sequence as the final window length, that is:

$$\min h_m^c = \frac{1}{(N-m+1)^2} \sum_{t_1=1}^{N-m+1} \sum_{t_2=1}^{N-m+1} p_{t_1}^c \ln\left(\frac{p_{t_1}^c}{p_{t_2}^c}\right) \tag{7}$$

$$\min h_m^r = \frac{1}{(N-m+1)^2} \sum_{t_1=1}^{N-m+1} \sum_{t_2=1}^{N-m+1} p_{t_1}^r \ln\left(\frac{p_{t_1}^r}{p_{t_2}^r}\right) \tag{8}$$

According to literature [13], the smaller the cross entropy, the smaller the difference between the two probability distributions, that is, the smaller the inter-group error between each window, the more stable the average absolute error fluctuation between each window, the more appropriate the number of corresponding window periods.

3.2. Model Introduction

Considering the center and range series of interval data, a partially additive linear model with mid-point range is obtained:

$$Y_t^c = X^c\beta^c + \sum_{i=1}^{m} g_i^c(Z_{it}^c) + \varepsilon_t^c, t = 1,2,\ldots,T \quad (9)$$

$$Y_t^r = X^r\beta^r + \sum_{i=1}^{m} g_i^r(Z_{it}^r) + \varepsilon_t^r, t = 1,2,\ldots,T \quad (10)$$

Where, $Y_t^c$and Y_t^rrepresent the center and extreme difference columns Y_t of the interval time series, X^c and X^r represent the center and range series vectors of the vector interval time series X of the linear explanatory variables, $g_i^c$and $g_i^r$are unknown smooth functions, Z_{it}^c and Z_{it}^r denote the center and range series of interval time series Z_{it} of explanatory variables modeled in a nonparametric manner; $\varepsilon_t^c$and ε_t^r are additive new interest terms [14], and satisfies the $E[\varepsilon_t^c \backslash I_{t-1}] = 0$ and $E[\varepsilon_t^r \backslash I_{t-1}] = 0$.

Model (3) and model (4) are denoted as C-PALM and R-PALM respectively. Using C-PALM as an example, it can be seen that the partially additive linear model has more flexibility than the general linear model. The right side of equation (3) is composed of two parts: one is the linear main part $X^c\beta$, which is a linear combination of covariables X^c, representing the linear relationship between Y and X, grasping the trend of the general trend, and is used for epitaxial prediction; The second is the non-parametric part $g_i^c(.)$,which is the non-parametric function of the covariate Z^c, representing the unknown functional relationship between Y and Z, which is used as a local adjustment to make the fitted data more accurate.

3.3. Parameter estimation

In this paper, we choose to use weight function to estimate unknown nonlinear regression function. There are two ways to solve this kind of estimation problem [15]: one is to combine the weight function with the least square method. In the selection of weight, some adjustments are made by the least square method; The second is to decompose the model into linear and nonlinear parts, using the least square method to estimate the former, using the weight function to estimate the latter. Since multivariate data are used in the empirical analysis in this paper, Scheme 2 is adopted. The basic idea of scheme 2 is: when the best window period number is m, first set the initial value $\beta^{c(0)}$, then kernel function estimation method is used to kernel estimate m nonlinear regression functions, and then the final β^c is obtained by updating β^c based on the least square method. The specific iterative algorithm steps are as follows:

Step 1: Given the initial value $\beta^{c(0)} = 0, g_i{}^{c(0)}(z) = 0$,i=1, 2,...,m, let k = 0.

Step 2: Let $k = k + 1$, $s = 1,2,\ldots,m$ do the following iterations in turn:

① Calculate the residual: $e_s^{c(k)} = y_t^c - \beta^{c(k-1)}X^c - \sum_{i=1}^{s-1} g_i^{c(k)}(z_{it}^c) - \sum_{i=s+1}^{m} g_i^{c(k-1)}(z_{it}^c)$, where c represents the central sequence and k represents the number of iterations.

② Perform kernel estimation: $g_s^{c(k)}(z^c) = \sum_{t=1}^{n} W_{nt}^c(z^c)e_s^{c(k)}$, where $W_{nt}^c(z) = K(\frac{z_t^c - z^c}{h_j}) / \sum_{t=1}^{n} K(\frac{z_t^c - z^c}{h_j})$ is the weight function, $K(.)$ is the verification function Number, h_j is the window width.

Step 3: Update β^c based on least square method. In the case of estimating m unknown functions $g_i^c(z_{it}^c)$ with kernel function, we get:

$$Y_t^c - \sum_{t=1}^{m} g_i^{c(k)}(z_{it}^c) = X^c\beta^{c(k)} + \varepsilon_t^c \quad (11)$$

Let$Y_t^{c(k)} = Y_t^c - \sum_{t=1}^m g_i^{c(k)}(z_{it}^c)$, then has $Y_t^{c(k)} = X^c\beta^{c(k)} + \varepsilon_t^c$. The sample model has changed from a formal to a conventional multivariate linear model, then we can use the least squares method to estimate β^c:

$$\widetilde{\beta^{c(k)}} = [(X^c)'X^c]^{-1}(X^c)'\,Y^{c(k)} \tag{12}$$

Step 4: Repeat steps 2 and 3 until the difference between the sum of squares of residuals after the k update and the sum of squares of residuals after the k+1 update is less than a set threshold and the update stops. To wit:

$$e = \sum_{j=1}^n (Y_t^c - \beta^{c(k)}X^c - \sum_{t=1}^m g_i^{c(k)}(z_{it}^c))^2 - \sum_{j=1}^n (Y_t^c - \beta^{c(k+1)}X^c - \sum_{t=1}^m g_i^{c(k+1)}(z_{it}^c))^2 \tag{13}$$

Suppose the threshold is ω, that is, updates stop when $e \le \omega$.

Based on the above algorithm, estimates of β^c, β^r, g_i^c and g_i^r can be obtained. Therefore, when the optimal number of window periods is m, a fixed weight is assigned to the sub-prediction sequence of center and range obtained by the partially additive linear model based on the sliding window, that is, $1/n$ of the total number of repeated prediction moments in the sub-prediction sequence.

Among them, the choice of the window width of kernel estimation determines the accuracy of estimation and the convergence speed of algorithm. The commonly used window width selection methods include experience selection method, staggered identification selection method, minimizing integral mean square error method, and theoretical window width optimal selection method, etc. The window width selected for this article is $h_j = c_j\sigma_j n^{-1/(2P+l)}$, where, c_j is the constant to be estimated that depends only on the density function and kernel function of the explanatory variables in the regression and is independent of n, σ_j is the standard deviation of the continuous variable j, n is the number of observed data, P is the order of the kernel function, and l is the number of linear continuous variables. Choose the Gauss kernel as the kernel function. In the judgment of iterative convergence of the algorithm, the threshold value ω is selected as 10-8. According to theorem 4.1.3 in reference [16], the function $\widehat{g_i^c}$ and $\widehat{g_i^r}$ have strong consistency, and according to the correlation proof in reference [17], the estimator $\widehat{\beta^c}$ and $\widehat{\beta^r}$ have strong consistency.

3.4. Prediction error measure of interval valued time series

In order to intuitively see the prediction effect of the midpoint range partial additive model based on the sliding window, this paper introduces several common interval valued time sequence prediction error measurement indicators [10], as shown in Table 1.

Table 1. Index of forecasting error measurement of interval valued time series

Measures	Measure formula	Pros and cons
Mean interval position error sum of squares(MSEP)	$\frac{1}{n}\sum_{t=1}^{n}(c_t-\hat{c_t})^2$	Different intervals can have the same center and need to be combined with the radius sequence error measure
Mean interval length error sum of squares(MSEL)	$\frac{1}{n}\sum_{t=1}^{n}(r_t-\hat{r_t})^2$	Different intervals can have the same radius and need to be combined with the central sequence error measure
Mean interval error sum of squares (MSEI)	$\frac{1}{n}\sum_{t=1}^{n}(c_t-\hat{c_t})^2$	Absolute indicators, which represent errors in the form of sum of squares, are susceptible to dimensional and extreme values
Mean relative interval error sum (MRIE)	$+\frac{1}{n}\sum_{t=1}^{n}(r_t-\hat{r_t})^2$	The center of the predicted interval and the actual random interval are identical to 0, and the influence of radius is ignored
	$\frac{1}{n}\sum_{t=1}^{n}\frac{\setminus c_t-\hat{c_t}\setminus}{r_t+\hat{r_t}}$	

4. Empirical analysis

In this paper, the quarterly growth rate of GDP is selected as a metric, and four financial variables (stock market volatility, money supply (M1), term structure spread and real effective exchange rate) are selected as predictive factors based on the results of economic and financial theory's economic boom analysis and research, which are closely related to the "troika" driving economic growth. In terms of data processing, this paper selects the maximum and minimum values of these economic and financial indicators in a certain year according to different data frequency, so as to construct the interval value time series of the corresponding indicators.

4.1. Specific description of the indicators

The specific description of relevant data in this paper is shown in Table 2.

Table 2. Description of indicators

Name of indicator	Frequency	Indicator Description	Data source
Gross domestic product	Quarterly	Quarter-over-quarter growth rate	Wind database
Stock market volatility (Shanghai Composite Index)	Daily	The monthly range is obtained by summating the daily range, and the monthly year-on-year change rate is calculated	Wind database
Money supply (M1)	Monthly	Monthly year-on-year growth rate	Wind database
Term structure spread	Monthly	Take the arithmetic average of the 5-year deposit rate and the 5-10 year loan rate as the long-term interest rate, and subtract the 3-month interbank lending rate from the long-term interest rate, and the difference is the term structure spread	China economic database
RMB real effective exchange rate	Monthly	Month-over-month rate of change	Bank for International Settlements (IBS) official website

Note: Sample period covers 1996-2019; GDP is denoted as Y, Shanghai Composite Index is denoted as $X1(Z_{1t})$, money supply $(M1)$ is denoted as $X2(Z_{2t})$, term structure spread is denoted as $X3(Z_{3t})$, and RMB real effective exchange rate is denoted as $X4(Z_{4t})$.

4.2. Stationarity test

In order to ensure the stationarity of interval time series variables, ADF stationarity test was conducted on interval data center and range series of each variable. The specific test results are shown in Table 3.

Table 3. Shows the stationarity test results of interval valued time series variables

Name of indicator	Data sample	Center sequence		Range sequence		Test results
		ADF Test statistics	P values	ADF test statistic	P values	
Gross domestic product	Original interval sequence	-0.883	0.323	- 1.773	0.073	Uneven
	Hukuhara difference sequence	-5.537	0.000	-5.314	0.000	smooth
Stock market volatility (Shanghai Composite Index)	Original interval sequence	-2.607	0.086	- 1.633	0.095	Uneven
	Hukuhara difference sequence	-6.347	0.000	-6.906	0.000	smooth
Money supply (M1)	Original interval sequence	-0.881	0.321	- 1.451	0. 133	Uneven
	Hukuhara difference sequence	-3.668	0.001	-4.522	0.000	smooth
Term structure spread	Original interval sequence	- 1.524	0. 117	- 1.249	0. 188	Uneven
	Hukuhara difference sequence	-6.409	0.000	-6. 140	0.000	smooth
RMB real effective exchange rate	Original interval sequence	-3.504	0.063	- 1.319	0. 167	Uneven
	Hukuhara difference sequence	-7.670	0.000	-6.282	0.0000	smooth

As can be seen from the test results in Table 3, the central and range series of GDP, stock market volatility, money supply (M1), term structure spread and RMB real effective exchange rate are not stable, but the series after Hukuhara difference is stable. Therefore, the central and range series of the explanatory variables and the explained variables after Hukuhara difference are used in the prediction model.

4.3. The midpoint range part can be added to the order of the linear model

In order to see the correlation between the interval valued time series variable GDP and the stock market volatility, the money supply (M1) term structure spread and the real effective exchange rate of RMB, the data were standardized and the scatter plot matrix of the center and range series was obtained. After sorting, the results were shown in Table 4.

Table 4. Correlation between macroeconomic indicators and various financial indicators

Correlation coefficient	Stock market volatility	Money supply (M1)	Term structure spread	RMB real effective exchange rate
Central sequence	-0.267	0.477	0.468	0. 112
Range sequence	0.069	0.639	0.022	0.402

According to the mean square error results in Table 5, the optimal model for both center and range series is to select a linear explanatory variable.

Table 5. lation the central range can be added to the order of the linear model

	Model form	Linear variable selection	Mean square error (MSE)
Central sequence	$Y_t = \beta_1 X_1 + \beta_1 X_2 + \sum_{i=1}^{2} g_i(Z_{it})$	Money supply (M1) term structure spread	2.076
	$Y_t = \beta_1 X_1 + \sum_{i=1}^{3} g_i(Z_{it})$	Money supply (M1)	0.171
Range sequence	$Y_t = \beta_1 X_1 + \beta_1 X_2 + \sum_{i=1}^{2} g_i(Z_{it})$	Money supply (M1) RMB real effective exchange rate	0.083
	$Y_t = \beta_1 X_1 + \sum_{i=1}^{3} g_i(Z_{it})$	Money supply (M1)	0.041

4.4. Determination of the number of sliding window periods

Based on the partially additive linear model of midpoint range, the sliding window model was applied to the center and range sequence after Huku-hara difference, and the optimal number of window periods was determined based on the cross-entropy principle. The results are shown in Table 6 and Table 7.

Table 6. able for determining the number of window periods in the center sequence

Number of window periods	2	3	4	…	20	21	22
Cross entropy	0.168	0.167	0.150	…	0.055	0.053	0.050

Table 7. Determination table of the number of window periods of range sequence

Number of window periods	2	3	4	…	20	21	22
Cross entropy	0.179	0.178	0.156	…	0.059	0.057	0.054

For center and range sequences, a smaller cross entropy during the window indicates a smaller difference between the two probability distributions, so the optimal number of window periods is 22 for both.

4.5. Prediction Results

Three models, vector error correction model (VECM), partial linear model (PLM) and text model (PALM), were used for modeling, and the parameter estimation results of the center and range sequence obtained were shown in Table 8.

Table 8. Parameter estimation results of the three models

model	C-VECM		R-VECM	
Parameter estimates	β0 = -0. 154, β1 = 0. 144, β2 = 0.028, β3 = 0.023 β4 = 0.669, β5 = -0.559		β0 = 0.097, β1 = -0.255, β2 = -0.002, β3 = 0.028, β4 = 0.073, β5 = 0.277	
model	C-PLM	R-PLM	C-PALM	R-PALM
Parameter estimates	β1 = -0.0004 β2 = 0. 106 β3 = 0.761	β1 = -0.003 β2 = 0.081 β3 = 0. 139	β1 = 0.014 β2 = 0.034	β1 = 0.008 β2 = 0.019

Note: After testing, the optimal lag order of center and range sequence of VECM model is 2; The parameter estimates obtained by the model in this paper are different at different times, and only the parameter estimates of the first and last sequence are listed.

Figure 2 shows the distribution of prediction intervals of the original interval and the vector error correction model, where the solid line represents the original interval value and the dashed line represents the predicted interval value. Figure 3 and Figure 4 show the distribution of the original interval and the partial linear model and the prediction interval of the model in this paper respectively. It can be seen that the prediction interval in FIG. 2 can generally reflect the overall trend, the prediction interval in FIG. 3 is generally small, and the prediction interval in FIG. 4 can more accurately fit the original interval. Table 9 lists the forecast intervals and original interval values of the macroeconomic metric GDP under the three regression model forecasting methods.

Table 9. Comparison of the original interval with the predicted interval

Time	Original interval		Vector error correction model		Partially linear model		Textual model	
1998	[7. 1,7.8]	(7.45,0.35)	[8.20,9.58]	(8.89,0.69)	[7.64,8.69]	(8. 16,0.52)	[7.08,7.78]	(7.43,0.35)
1999	[7.7,8.9]	(8.3,0.6)	[7.62,8.52]	(8.07,0.45)	[7.90,8.43]	(8. 17,0.26)	[7.70,8.90]	(8.30,0.60)
2000	[8.5,8.9]	(8.7,0.2)	[7.40,7.62]	(7.51,011)	[8.87,9.67]	(9.27,0.40)	[8.29,8.70]	(8.49,0.20)
…	…	…	…	…	…	…	…	…
2017	[6.3,6.4]	(6.95,0.05)	[7.59,7.62]	(7.61,-0.02)	[8.36,8.95]	(8.65,0.29)	[6.84,7. 10]	(6.97,0. 13)
2018	[6. 1,6.2]	(6.8,0. 1)	[6.32,6.46]	(6.39,0.07)	[7.32,8.62]	(7.97,0.65)	[6.74,7.26]	(7.00,0.26)
2019	[5.96,6.0]	(6.2,0.2)	[6.32,7. 19]	(6.75,0.43)	[7.73,8.09]	(7.91,0. 18)	[5.99,6.43]	(6.21,0.22)

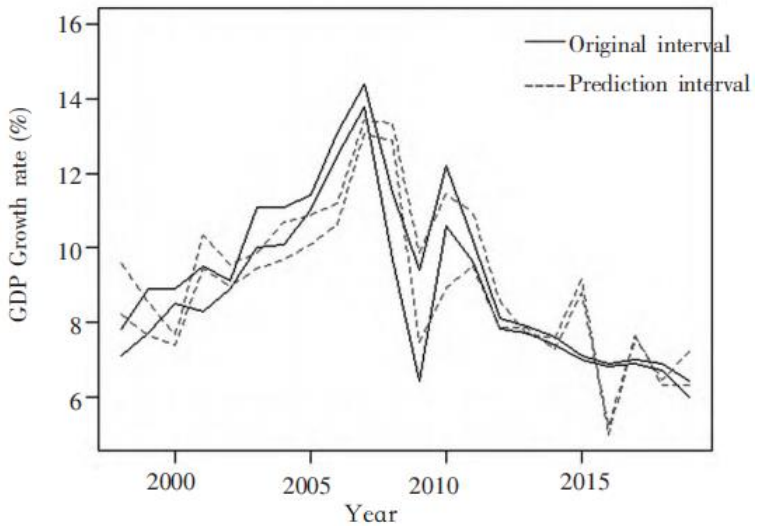

Figure 2. Original interval and vector error correction model prediction interval

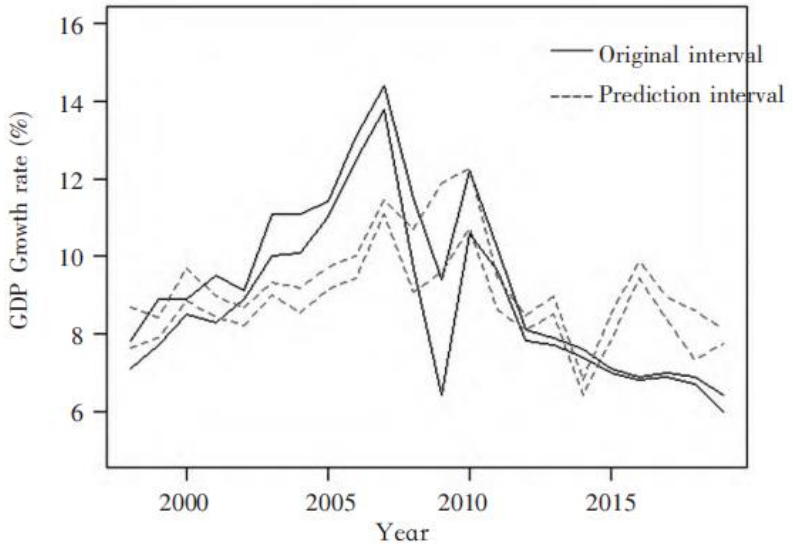

Figure 3. Original interval and partial linear model prediction interval

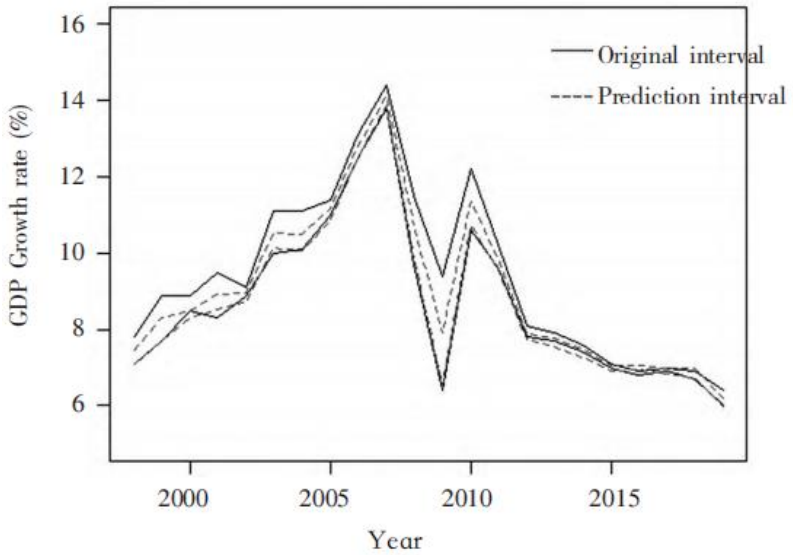

Figure 4. Original interval and prediction interval of the model in this paper

As can be seen from the prediction effect of Table 10, the values of multiple indexes of some linear models are the largest, followed by the vector error correction model, while the values of each index of the partially additive linear model in this paper are small, indicating that it has achieved a better fitting effect. It can be seen that after the introduction of multiple non-parametric smooth functions, the fitting of the data becomes more refined, and the model in this paper is superior to these traditional regression models.

Table 10. Comparison of prediction effect

model	MSEP	MSEL	MSEI	MRIE
Vector error correction model	1.209	0.068	1.277	1.368
Partially linear model	2.458	0.052	2.510	1.669
Textual model	0.008	0.011	0.019	0. 154

5. Conclusion

Based on the partial linear model, this paper combined the additive model and introduced the sliding window model. For the interval type data, a partially additive linear model based on the sliding window was proposed, which combined the advantages of semi-parametric regression and the sliding

window model, and avoided the dimension disaster, and was more precise in data fitting. On the selection of sliding window periods, the cross-entropy criterion is introduced to determine the best window periods. In parameter estimation, a two-step iterative algorithm, least square method and kernel estimation method are used to estimate the unknown parameters and unknown functions of the model under the optimal number of window periods. In the empirical analysis part, based on three similar models, some financial indicators are introduced to forecast and analyze the macro economy. The results show that the model has a better fitting effect on the macroeconomic forecast under the interval data. In the future research, we can further consider adding autoregressive term to the model, and compare the influence of different window width selection methods and different kernel function selection on the estimation results.

References

[1] Haohui Ma and Xuemei Sun. 2023. The Design of Enterprise Management Information System Based on Cloud Computing. In Proceedings of the 2023 6th International Conference on Software Engineering and Information Management (ICSIM '23). Association for Computing Machinery, New York, NY, USA, 81–85. https://doi.org/10.1145/3584871.3584883.

[2] A. Bochkarev, A. Urasova, and D. A. Balandin. 2022. Methodological aspects of information support in the enterprise management system. In IV International Scientific and Practical Conference (DEFIN-2021). Association for Computing Machinery, New York, NY, USA, Article 17, 1 – 4. https://doi.org/10.1145/3487757.3490853.

[3] Zhiyi Zhuo and Shanhu Zhang. 2019. Research on the Application of Big Data Management in Enterprise Management Decision-making and Execution Literature Review. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing (ICMLC '19). Association for Computing Machinery, New York, NY, USA, 268 – 273. https://doi.org/10.1145/3318299.3318388.

[4] Shufang Sun. 2021. Analysis of Effective Ways of Human Resource Management in Enterprise Management Based on Risk Early Warning Model. In 2021 4th International Conference on Information Systems and Computer Aided Education (ICISCAE 2021). Association for Computing Machinery, New York, NY, USA, 2343 – 2347. https://doi.org/10.1145/3482632.3487426.

[5] David Sundaram and Elke Wolf. 2009. Enterprise model management systems. In Proceedings of the International Workshop on Enterprises & Organizational Modeling and Simulation (EOMAS '09). Association for Computing Machinery, New York, NY, USA, Article 5, 1 – 15. https://doi.org/10.1145/1750405.1750412.

[6] Meiryani Meiryani, Michael Evan Angeleus, and et al. 2023. Impact of Enterprise Risk Management Implementation on Fraud Control in Small and Medium Enterprises. In Proceedings of the 2022 6th International Conference on E-Business and Internet (ICEBI '22). Association for Computing Machinery, New York, NY, USA, 340 – 349. https://doi.org/10.1145/3572647.3572698.

[7] Guannan Wang, Wenjia Wang, Xueliang Song, and Jiayi Chen. 2022. Machine Learning-based Pipeline for Enterprise Cross-department Conflict Data Management. In 2021 4th International Conference on Computing and Big Data (ICCBD 2021). Association for Computing Machinery, New York, NY, USA, 33 – 36. https://doi.org/10.1145/3507524.3507530.

[8] Arun Kumar, Matthias Boehm, and Jun Yang. 2017. Data Management in Machine Learning: Challenges, Techniques, and Systems. In Proceedings of the 2017 ACM International Conference on Management of Data (SIGMOD '17). Association for Computing Machinery, New York, NY, USA, 1717 – 1722. https://doi.org/10.1145/3035918.3054775.

[9] Zhiyi Zhuo and Shanhu Zhang. 2019. Research on the Application of Big Data Management in Enterprise Management Decision-making and Execution Literature Review. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing (ICMLC '19). Association for Computing Machinery, New York, NY, USA, 268 – 273. https://doi.org/10.1145/3318299.3318388.

[10] Pratyusa K. Manadhata. 2015. Machine Learning for Enterprise Security. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (AISec '15). Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/2808769.2808782.

[11] Zhihan Lv, Ranran Lou, Hailin Feng, Dongliang Chen, and Haibin Lv. 2021. Novel Machine Learning for Big Data Analytics in Intelligent Support Information Management Systems. ACM Trans. Manage. Inf. Syst. 13, 1, Article 7 (March 2022), 21 pages. https://doi.org/10.1145/3469890.

[12] Junhong He and Ke Ma. 2021. Enterprise Financial Risk Management and Control. In 2021 2nd Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC2021). Association for Computing Machinery, New York, NY, USA, 393 – 396. https://doi.org/10.1145/3452446.3452547.