Research Article

Research on Small Sample Nonlinear Cointegration Test and Modeling Based on the LS-SVM Optimized by PSO

Jungang Du

Department of Management Engineering and Equipment Economy, Naval University of Engineering, Wuhan 430033, China

Correspondence should be addressed to Jungang Du; 398255218@qq.com

Received 19 April 2022; Revised 5 August 2022; Accepted 8 August 2022; Published 26 August 2022

Copyright © 2022 Jungang Du. 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.

According to the definition of nonlinear cointegration, this article studies the small sample nonlinear cointegration test and NECM (Nonlinear Error Correction Model) based on the LS-SVM (Least Squares Support Vector Machine) optimized by PSO (Particle Swarm Optimization). And the logical process of this method is also designed. Then, we carry out empirical research on the ship maintenance cost index and the several price indexes. Based on the judgment of the type of cointegration test relationship, the test of the nonlinear cointegration test relationship of small samples is realized, and the NECM for predicting the ship maintenance cost index is established as well, which is compared with the VAR model of linear vector. The research result shows that the small sample nonlinear cointegration test and modeling method based on the LS-SVM Optimized by PSO can describe the nonlinear cointegration test relationship of the small sample system. And the NECM has better performance. The prediction effect can effectively predict small sample nonlinear systems. We also compare the prediction results with the wavelet neural network algorithm, and the results show that the generalization ability of LS-SVM Optimized by PSO is better, and the prediction accuracy of small samples is higher.

1. Introduction

The modeling method for traditional econometrics is based on the asymptotic theory of least squares estimation, which premised on the stability of the time series. In economic systems, especially in macroeconomic, variables are mostly nonstationary. Therefore, traditional statistical inference cannot be directly applied. As a result, Engle and Granger have proposed the cointegration test theory [1] and pointed out that the cointegration system has three main forms, namely, the VAR (Vector Auto Regressive) model, the VMA (Vector Moving Average) model, and the ECM (Error Correction Model), which provide an approach to model nonstationary time series. Cointegration theory describes the long-term linear equilibrium relationship between each component sequence of the economic system and can be expressed by the ECM. Due to the pseudoregression phenomenon can be effectively avoided, the ECM based on linear cointegration technology was widely used in economic forecasting [2–4].

Linear cointegration implies two basic conditions. The first one is that each time series has an integer dimension, and the second one is that it has the same single integer order, which is the basis for the study of linear cointegration theory. However, many time series in the economic system have the characteristics of nonlinearity and long memory [5]. A single time series is not a single integer dimension but a fractal dimension or different integer dimensions, which means that the single time series is nonlinear and featured with fractal character, and the equilibrium relationship between sequences is often nonlinear as well. For nonlinear systems, the linear cointegration theory is not applicable, and the nonlinear cointegration theory shall be used to study the nonlinear equilibrium relationship between time series. Literature [6] gave the definition of nonlinear cointegration test:
As for a vector time series $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})^T$, the component sequence of $\{X_t\}$ shall be named as nonlinearly cointegrated, if

1. $x_t = (i = 1, \ldots, n)$ is a LMM (Long Memory in Mean, LMM) sequence or an I(1) sequence.

2. There is a nonlinear function $f()$, let $y_t = f(x_{1t}, x_{2t}, \ldots, x_{nt})$, which is an SMM (Short Memory in Mean, SMM) sequence.

The function $f()$ is named as the nonlinear cointegrating function. If the nonlinear cointegration function has linear properties with respect to the independent variables, that is

$$f(x_{1t}, x_{2t}, \ldots, x_{nt}) = \alpha_1 x_{1t} + \alpha_2 x_{2t} + \cdots + \alpha_n x_{nt},$$  \hspace{1cm} (1)$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the vector in $R^n$, and the long memory of the component sequence in $\{X_t\}$ is expressed as the I(1) sequence; the nonlinear cointegration shall be simplified to the linear cointegration relationship. Therefore, the definition of nonlinear cointegration is a generalization of the concept of linear cointegration, and the nonlinear cointegration is a special case of nonlinear cointegration.

As for the vector time series $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})^T$, if the component sequences are all LMM sequences, ($I(1)$ sequences are all LMM sequences; if the LMM sequence is $I(1)$, its single integer order is not necessarily equal.), and there is a nonlinear function $f()$ which ensure the $f(x_{1t}, x_{2t}, \ldots, x_{nt})$ is an SMM sequence. Therefore, $\{X_t\}$ is named as nonlinear cointegration.

The key element to the testing and modeling of nonlinear cointegration relationship lies in the estimation of the nonlinear cointegration function $f()$. In view of the fact that the general linear cointegration analysis method cannot realize the nonlinear cointegration test and modeling, the literature [6] proposed a neural network algorithm suitable for the test and modeling of the nonlinear cointegration relationship of vector time series and discussed the theoretical basis of the method and its feasibility. By using a wavelet neural network with the ability to approximate nonlinear functions arbitrarily, literature [7] studied the NECM of nonlinear cointegration systems, provided the modeling method, and carried out relevant analysis. These research found that compared with the linear VAR model, the NECM has a better prediction effect and can effectively predict the nonlinear economic system.

For the nonlinear system modeling problems, many meta-heuristic algorithms have emerged in recent years, such as the Monarch Butterfly Optimization (MBO) [8], Slime Mould Algorithm (SMA) [9], Moth Search Algorithm (MSA) [10], Hunter Games Search (HGS) [11], Runge Kutta Method (RUN) [12], Colony Prediction Algorithm (CPA) [13], and Harris Hawks Optimization (HHO) [14]. These algorithms have the characteristics of simple principle and easy implementation and have excellent performance in solving the optimization problems of complex nonlinear systems in the case of large samples. But the basic model of algorithm also has obvious defects; there is a lot of room for improvement.

For example, MBO has a single local search ability and limited search location, which will lead to the degradation of the population. Its global search method is too simple to make full use of the effective information of the population. However, adopting the elite retention strategy requires setting parameters and sorting operations, which increases the complexity of the algorithm [15]. The structure of SMA is simple and clear, but it is found that this algorithm still has some shortcomings, such as easy to fall into local optimization, slow convergence speed, low accuracy, and so on [16]. The MSA has the advantages of simple structure, few parameters, high accuracy, and strong robustness, but its development ability still needs to be improved, and it is easy to fall into local optimization [17]. In numerical analysis, the RUN is an important family of implicit and explicit iterative methods for the approximation of solutions of ordinary differential equations. But compared with other intelligent algorithms, the calculation process of this method is more complex. The CPA algorithm has long running time, intermittent ends, and difficult parameters to adjust [18]. For HHO algorithm, how to balance contradictions between the exploration and exploitation capabilities and alleviate the premature convergence are two critical concerns that need to be dealt with in the HHO study [19]. Like the above algorithms, HGS requires a high number of training samples and has a good application effect in large sample research objects, but it is not suitable for small samples [20].

The above research provides effective methods and technical tools for the testing and modeling of nonlinear cointegration relationships. However, these methods have certain requirements for the quantity and quality of each series, which are suitable for large sample time series (samples shall be larger than 100 pieces). The neural network algorithm and wavelet neural network shall be prone to overfitting phenomenon, and the generalization ability of the model is not good as well [21], including relational testing and modeling have poor applicability. In view of this, this article proposes to apply the LS-SVM [22] with excellent learning performance and strong generalization ability to test the nonlinear cointegration relationship among small sample time series. Meanwhile, the PSO, which is widely used in function optimization, neural network training, fuzzy system control, and other genetic algorithms, is selected to jointly optimize the parameters of LS-SVM. This article further studies the modeling of small-sample NECM by the LS-SVM optimized by PSO. According to the method proposed in this article, the nonlinear cointegration relationship among the ship maintenance cost index and several price indexes is tested and modeled, and the empirical analysis is carried out and compared with the VAR model and the wavelet neural network algorithm.

2. Nonlinear Cointegration Test and NECM Model

2.1. Nonlinear Cointegration Test. According to the definition of nonlinear cointegration, if each component of vector
time series $X_t = (x_{1t}, x_{2t}, \ldots, x_{mt})^T$ is LMM sequence, the SMM sequence with a constant mean value shall be used as the target value, and after conversion of nonlinear function $f()$, we may have the output sequence $y_t = f(x_{1t}, x_{2t}, \ldots, x_{mt})$. However, it cannot be confirmed that this $f()$ is an estimation of the nonlinear cointegration function, and it is necessary to check whether the residual sequence $\{u_t\}$ is an SMM sequence. If the sequence $\{u_t\}$ is an SMM sequence, then $f()$ can be seen as the estimation of a nonlinear cointegration function, which shows that there is a nonlinear cointegration relationship among the component sequences of $\{X_t\}$. The inspection process for $\{u_t\}$ as follows:

2.1.1. Unit Root Test. Given the null hypothesis is $H_0$: there is a unit root in $\{u_t\}$; the alternative hypothesis is $H_1$: there is no unit root in $\{u_t\}$.

If the null hypothesis $H_0$ is accepted, then $\{u_t\}$ shall be a LMM sequence, which means that there is no nonlinear cointegration relationship among the component sequences of the $\{X_t\}$. If the alternative hypothesis $H_1$ is accepted, then $\{u_t\}$ does not have a unit root. However, it cannot be determined that it belongs to the SMM sequence, and the sequence memory test of $\{u_t\}$ is also required as well.

2.1.2. Sequence Memory Test. Calculate the auto-correlation function sequence $\{\rho_k\}$ of the sequence $\{u_t\}$, namely:

\[ \rho_k = \frac{Cov(u_t, u_{t-k})}{\sigma^2} \]

\( k = 1, 2, \ldots \) , \( \rho_k = \rho_k/\rho_0 \).

Time series memory shall be mainly judged from the convergence speed of auto-correlation function $\rho_k$. As for short-memory time series, the auto-correlation function $\rho_k$ shall be decreased along with a negative exponential speed as $k \rightarrow \infty$, that is, $\rho_k \sim c_0 e^{-k}$ (where, $a$ and $c$ are constants, and $a>1$ and $c>0$). For long memory time series, the auto-correlation function $\rho_k$ shall be decreased along with the speed of negative exponent as $k \rightarrow \infty$, that is, $\rho_k \sim c_k 2^{-d-1} e^{-k}$ (where, $a>1$), and $d<0.5$. Therefore, the faster the auto-correlation function of time series decreases, the more likely it is to be a short memory time series.

Quantitative methods for memory test of time series include R/S test [21], modified R/S test [22], KPSS test [23], and LM test [24].

2.2. NECM Model. Similar to the error correction model (ECM) derived from linear cointegration, for the vector time series $X_t = (x_{1t}, x_{2t}, \ldots, x_{mt})^T$, when there is a nonlinear cointegration relationship among its component series, the derived NECM shall be

\[ \Delta x_t = \gamma_0 + \sum_{j=1}^k \gamma_j \Delta x_{t-(j)} + \gamma_{ik} f(X_{t-k}; \theta) + \varepsilon_t, \]  

(2)

where $f()$ is a nonlinear cointegration function, $\theta$ is a parameter vector, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})^T$ is a random error sequence, $I_j$ (j = 0, 1, 2, ..., k) is the coefficient matrix of n×1, and $k$ is the lag intervals for endogenous.

Then, the NECM of the ordinal number $i$ component of $\{X_t\}$ is

\[ \Delta x_{ti} = \gamma_{0i} + \sum_{j=1}^k \gamma_{ji} \Delta x_{t-(j)} + \gamma_{ik} f(X_{t-k}; \theta) + \varepsilon_{ti}. \]  

(3)

Estimate each parameter in formula (3) and bring the parameter estimation result into the original formula, then the NECM corresponding to each component can be obtained. The model can be used to predict each component sequence, and the ordinal number $h$ step prediction model of the ordinal number $i$ component shall be

\[ \Delta x_{ti(t+h)} = \gamma_{0i} + \sum_{j=1}^k \gamma_{ji} \Delta x_{t-(h+j)} + \gamma_{ik} f(X_{t-k}; \theta) + \varepsilon_{ti}, \]  

\( i = 1, 2, \ldots, n. \)  

\[ x_{ti(t+h)} = x_{ti(t+h-j)} + \Delta x_{ti(t+h)}, \]  

\( i = 1, 2, \ldots, n. \)  

The estimation $f(X_{t-k}; \theta)$ of the nonlinear function in NECM is the equilibrium error of $\{X_t\}$ at time $t-k$, and its function is to correct the difference sequence value, which is the same as the function of the error correction term in linear cointegration. As $h$ is different, the estimation $f(X_{t+h}; \theta)$ of the nonlinear cointegration function is also different. Therefore, NECM’s predictions of variables at different time periods are independent of each other.

3. Parameter Optimization of LS-SVM Model Based on PSO

Support Vector Machine (SVM) is an effective nonlinear problem processing tool developed in recent years, which has great advantages for solving small sample, nonlinear, and high-dimensional problems [25]. LS-SVM [26] is an extension of the standard SVM, which has good robustness and requires fewer parameters to be optimized, which is widely applied [27].

3.1. LS-SVM Model and Selection of Kernel Function. LS-SVM is obtained by transforming inequality constraints into equality constraints in standard SVM algorithm [28], which is a form of SVM under quadratic loss function.

In LS-SVM modeling, the role of the kernel function is equivalent to projecting the samples into a high-dimensional space, transforming it into a linear regression problem, and then constructing the optimal regression curve [29]. Therefore, the selection of the kernel function directly affects the generalization ability of the curve.

The common kernel function mainly includes polynomial kernel function $K(x, y) = (x^T y)^d$, sigmoid kernel function $K(x, y) = \tanh(k(x^T y - \theta))$, spline kernel function $K(x, x^T) = 1 + \exp(-\|x - y\|_2^2/2\sigma^2)$, and radial basis function (RBF) kernel function $K(x, y) = \exp(-\|x - y\|_2^2/2\sigma^2)$. The selection of the kernel function includes the determination of the form of the kernel function and the determination of the kernel parameters. Considering the strong generalization ability of the RBF
kernel function [30, 31], this article selects the RBF kernel function for modeling research.

3.2. Optimization of LS-SVM Model Parameters Based on PSO. After selecting RBF as the LS-SVM kernel function, there are 2 parameters shall be defined, namely the regularization parameter $\gamma$ and the kernel parameter $\sigma^2$. These two parameters largely determine the learning ability and prediction ability of LS-SVM. Therefore, it is necessary to find the optimal combination of $\gamma$ and $\sigma^2$.

At present, for the optimal value of LS-SVM parameters, the general selection method is grid search method, which achieves satisfactory results through continuous experiments. However, this method is time-consuming and inefficient. Some scholars proposed to use the gradient descent method to select the parameters of LS-SVM [32]. However, this method is limited by the fact that the kernel function must be differentiable, and it is easy to fall into local minima in the search process. Some scholars have also tried to use immune algorithm to optimize the parameters of LS-SVM [33] to reduce the blindness of parameter selection and improve the prediction accuracy of LS-SVM, but the implementation of this method is complicated. Besides, some scholars proposed to use genetic algorithm to determine LS-SVM parameters [34], but genetic algorithm needs to perform crossover and mutation operations, and many parameters need to be adjusted, which is computationally complex and inefficient.

The PSO is a new stochastic optimization algorithm based on swarm intelligence. Similar to the genetic algorithm, PSO is also a random search optimization tool based on population, but there is no crossover and mutation operation. Particles follow the optimal particle in the solution space to search. It has the characteristics of parallel processing, good robustness, simple and easy to implement, high computational efficiency, and can find the global optimal solution of the problem with a large probability. Therefore, the PSO algorithm was selected in this article to optimize the parameters of LS-SVM, and the nonlinear cointegration relationship of small sample time series was tested based on the LS-SVM optimized by PSO.

3.3. PSO Algorithm. In the PSO algorithm, each alternative solution is called a “Particle”. Multiple particles coexist and cooperatively search for optimization. Each particle flies to a better position in the problem space and searches for the optimal solution according to its own “Experience” and the best “Experience” of the adjacent particle swarm. Each particle in the particle swarm represents a potential solution of the system. Each particle is represented by three indicators: position, speed, and fitness. The mathematical expression of PSO algorithm is as follows [35]:

Suppose in an $n$-dimensional search space, there are $m$ particles forming a group, where $X_i=(X_{i1}, X_{i2}, \ldots, X_{im})$ is the current position of particle $i$; $V_i=(V_{i1}, V_{i2}, \ldots, V_{im})$ is the current flight speed of particle $i$; $P_i=(P_{i1}, P_{i2}, \ldots, P_{im})$ is the position with the best fitness value experienced by particle $i$, which is called the individual optimal position; $P_g=(P_{g1}, P_{g2}, \ldots, P_{gm})$ represents the position of the optimal fitness value searched by all particles in the whole particle swarm so far, which is called the global optimal position. The position of each particle varies according to the following formula:

$$V_i(k+1) = w \cdot V_i(k) + c_1 \cdot \text{rand}_1 \cdot (P_i - X_i(k)) + c_2 \cdot \text{rand}_2 \cdot (P_g - X_i(k)),$$

$$X_i(k+1) = X_i(k) + V_i(k+1),$$

where $k$ represents the $k$th generation of evolution; $w$ is the inertia weight, indicating to what extent the original speed of the particle can be retained, the larger $w$ has better global convergence ability, while the smaller $w$ has a stronger local convergence ability, in order to slow down the movement process of the particles and prevent the oscillation phenomenon when the particles move towards $P_g$, generally, $w$ is defined linearly decrease with the evolution; $c_1$ and $c_2$ are two positive learning factors, and $c_1$ adjusts the step size of particles flying to the individual optimal position $P_i$, $c_2$ adjusts the step size of the particle flying to the global optimal position $P_g$. The larger the value, the greater the acceleration of the particle flying to $P_i$ and $P_g$. Generally, their value is 2.0. $\text{rand}_1()$ and $\text{rand}_2()$ are random numbers in the range [0,1].

Meanwhile, in order to reduce the possibility of particles leaving the search space during evolution, $V_i$ is usually limited to a certain range, namely $V_i \in [-V_{\text{max}}, V_{\text{max}}]$. $V_{\text{max}}$ is a constant value. In general, if the search space of the problem is limited to $[-X_{\text{max}}, X_{\text{max}}]$, then $V_{\text{max}} = \alpha \cdot X_{\text{max}}$ can be defined, generally $0.1 \leq \alpha \leq 1.0$. The number of particles $m$ is generally 20~40, the length $n$ is determined by the optimization problem, and the range of particles is also determined by the optimization problem, and different ranges can be set for each dimension. The termination condition (fitness) of the algorithm can be defined as the maximum number of cycles or the minimum error requirement, or it can be defined by the specific problem [36].

![Figure 1: Logic flow of LS-SVM optimized by PSO.](image)
3.4. Basic Procedures for LS-SVM Optimized by PSO. The LS-SVM kernel function selected in this article is RBF, and the parameters to be optimized are $\gamma$ and $\sigma^2$. Therefore, the particle length in PSO shall be defined as 2. The particle fitness shall select the mean square error (MSE) between the LS-SVM target value and the output value, whose function form is

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$  \hspace{1cm} (7)$$

where $n$ is the number of training samples, $y_i$ is the target value of LS-SVM, and $\hat{y}_i$ is the output value. The smaller the particle fitness means, the better current particle position.

According to the basic principle and steps of PSO algorithm, the logic flow for optimizing the parameters of $\gamma$ and $\sigma^2$ LS-SVM is shown in Figure 1. The optimization steps are as follows:

1. **PSO initialization**, define the particle swarm number as $m$, the maximum number of iterations as $T$, the initial value of inertia weight $w$ as 0.9, which decreased linearly with the increase of iteration times, $w(t) = 0.9 - t/T \ast 0.5$. Define the maximum speed $V_{\text{max}}$, random initialization of particle swarm in each particle position and velocity, namely $t = 0$, noted as $X_{i}(0) = (\gamma_{i}(0), \sigma^2_{i}(0))$, $V_{i}(0) = (V_{\gamma_{i}}(0), V_{\sigma^2_{i}}(0))$, $i = 1, 2, ..., m$.

2. Calculate the fitness of each particle. The values of $y_i$ and $\sigma^2$ contained in the current position of each particle shall be substituted into the LS-SVM model to obtain the model output value $\hat{y}_i$, which shall be compared with the target value $y_i$ to calculate the fitness of each particle.

For each particle, if the fitness (that is, the MSE of the LS-SVM target value and the output value) of its current position is smaller than the fitness of its previous best position $P_i$, the best position of this particle is its current position.

For each particle, if the fitness of its best position $P_i$ is smaller than the fitness of the global best position $P_g$ of all particles, the global best position $P_g$ will be replaced by $P_i$.

Change the velocity and position of the particle according to formula (5) and (6), so as to generate new particles.

Check whether the end condition is reached, if not, return to Step 2.

Obtain the optimal particle information, and assign the optimal combination of $\gamma$ and $\sigma^2$ to the LS-SVM model.

4. Nonlinear Cointegration Test and NECM Modeling Based on LS-SVM Optimized by PSO

The nonlinear cointegration relationship of small sample time series shall be tested by the method of LS-SVM optimized based on PSO; thus, the NECM model is established. The basic steps are as follows:

Check whether each component sequence has a linear cointegration relationship. The method is to calculate the fractional integration order of each sequence. If the fractional integration order is different, it means that each component sequence does not have a linear cointegration relationship.

Based on the LS-SVM optimized by PSO, the nonlinear cointegration relationship among small sample time series shall be tested, and the nonlinear cointegration function $f()$ shall be estimated and tested as well. The input of LS-SVM is each small sample sequence to be tested, and the target value is the SMM sequence with constant mean value. The minimization of the mean square error (MSE) between the target value and the output value is used as the training target, and the kernel function of LS-SVM is optimized based on PSO algorithm to obtain the optimization model.

The memory test is performed on the optimized output sequence. The memory test method adopts the modified R/S statistic $Q_n^*$ proposed by Lo [21], and the distribution function of $Q_n^*$ shall be

$$F(v) = 1 + 2 \sum_{k=1}^{\infty} (1 - 4k^2 v^2) e^{-2k^2 v^2},$$ \hspace{1cm} (8)$$

where $v$ is the quantile and $F(v)$ is the cumulative probability, that is, $F(v) = p(Q_n < v)$. The larger the $F(v)$, the greater the probability that the sequence belongs to the long memory sequence. When $v$ is 3, $F(v)$ reaches 0.9999. When the long memory is not significant, the output sequence of LS-SVM is a short memory time series, indicating that there is a nonlinear cointegration relationship between the sequences, and $f()$ is the nonlinear cointegration function of the sequence to be tested.

Substitute the obtained nonlinear cointegration function $f()$ into formula (3) and use the least squares method to estimate the parameters based on Eviews software.

Substitute the parameters into the prediction models (4) and (5), and the calculation result shall be the predicted value.

5. Empirical Analysis

Based on the 2000–2021 Ship Maintenance Price Index (SMPI), Consumer Price Index (CPI), Product Price Index (PPI), Materials, Fuel, and Power Purchasing Price Index (MPI) consists of 22 groups of data as samples for empirical analysis (The index in 2000 is 100), and the trend of each index is shown in Figure 2.

By using the LS-SVM optimized by PSO, we test the nonlinear cointegration relationship among SMPI, CPI, PPI,
and MPI, then establish the NECM prediction model of SMPI according to the nonlinear cointegration function \( f() \).

5.1. Judgment of the Type of Cointegration Relationship

5.1.1. Sequence Memory Test and Calculation Method of Integer Order. The basic calculation steps for using R/S test to judge sequence memory are as follows:

Define the time series \( \{x_t, t=1, 2, \ldots, n\} \), definition:

The sample mean value shall be
\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\] (9)

The sample variance shall be
\[
\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.
\] (10)

The accumulated deviation shall be
\[
y_i = \sum_{j=1}^{i} (x_j - \bar{x}).
\] (11)

The range value shall be
\[
R_n = \max_{1 \leq i \leq n} \sum_{j=1}^{i} (x_j - \bar{x}) - \min_{1 \leq i \leq n} \sum_{j=1}^{i} (x_j - \bar{x})
\] (12)

= \max_{1 \leq i \leq n} y_i - \min_{1 \leq i \leq n} y_i.

Then the estimated value of R/S statistic shall be
\[
Q_n = \frac{R}{S} = \frac{R_n}{\sigma_n}.
\] (13)

In case of short memory and heteroscedasticity, the R/S statistic is not robust. Lo [37] proposed a revised R/S statistic as
\[
Q_n^* = \frac{R}{S} = \frac{R_n}{\sigma_n^*}
\] (14)

where \( \sigma_n^2 = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^2 + (2/n) \sum_{i=1}^{q} \omega_i (q) [\sum_{k=i+1}^{n} (x_k - \bar{x})(x_{k-i} - \bar{x})] \), \( \omega_i (q) = 1 - i/q + 1, q < n \).

As for a time series \( \{x_t, t=1, 2, \ldots, n\} \), taking \( m \) observations of the series, the \( \{x_t, t=1, 2, \ldots, n\} \) is divided into \( l=\lceil n/m \rceil \) independent time series with length \( l \). Calculate the modified R/S test statistic \( Q_m^* \) shall be calculated to obtain the modified R/S test statistics of time series with length \( m \).
\[
Q_m = \frac{1}{l} \sum_{i=1}^{l} Q_i^m.
\] (15)

As for different \( m \) values, the modified R/S statistic sequence can be obtained, and Mandelbrot proved that
\[
P \lim_{m \to \infty} \left( m^{-H} Q_m \right) = C,
\] (16)

where \( C \) is a constant, and \( H \) is the Hurst exponent. That is: \( Q_m \propto m^{H} \), taking the logarithm of both sides, we may have
\[
\ln Q_m = \ln C + H \ln m.
\] (17)

By using least squares regression, \( H \) can be obtained. The relationship between \( d \) and \( H \) shall be \( H = d + 1/2 \). Thus, an estimated value of the fractional order \( d \) is obtained.

5.1.2. Cointegration Relationship Types among SMPI, CPI, PPI, and MPI. According to the above calculation methods...
of sequence memory and fractional order, the memory of SMPI, CPI, PPI, and MPI in the sample interval is tested. The modified R/S statistic $Q_n^*$ of each index is shown in Table 1.

From the test results in Table 1, the modified R/S statistic $Q_n^*$ of SMPI, CPI, PPI, and MPI is far greater than 3. These time series have significant long memory characteristics and belong to long memory time series. The following is the calculation of the fractional order of each index, taking SMPI as an example.

If we define value $m$ as 1, 2, 3, 4, and 5, respectively, for finding out the statistic $Q_n^*$ corresponding to SMPI, the results are shown in Table 2.

The modified R/S statistic $Q_n^*$ and the corresponding $m$ value of SMPI are substituted into Equation (17). $H = 0.804$ was obtained by least square regression. Therefore, the fractional order of SMPI is $d = H - 0.5 = 0.304$.

By using the same method and steps, the fractional order of CPI, PPI, and MPI shall be calculated separately. The results are shown in Table 3.

According to the statistic $Q_n^*$ test results of SMPI, CPI, PPI, and MPI, and the calculation of the fractional order of each index, we find out that each index belongs to the long memory time series, and the fractional order of each index is not the same. Therefore, there is no linear cointegration relationship among SMPI, CPI, PPI, and MPI. Whether there is a nonlinear cointegration relationship among them requires further testing.

5.2. Estimation and Testing of Nonlinear Cointegration Relationships. According to the steps of nonlinear cointegration relationship test described in 4.2, this article selects 18 groups of data (SMPI, CPI, PPI, and MPI) from 2000 to 2017 as training samples to estimate the nonlinear function relationship and optimize the LS-SVM parameters. Then, the 4 groups of data (SMPI, CPI, PPI, and MPI) from 2018 to 2021 are used as test samples and substituted into the optimized LS-SVM model by rolling verification. The revised R/S statistic $Q_n^*$ are estimated for 19, 20, 21, and 22 output sequences to test whether they are short memory sequences.

In the training and verification of LS-SVM model, the input variables are SMPI, CPI, PPI, and MPI, and their minimum value is 100 (that is, the benchmark value in 2000). Among these variables, SMPI has the largest value, but its maximum value is less than 600. Considering the possibility of prediction range, this article sets the maximum value of input variables as 1000. Therefore, the range of input variables of LS-SVM model is [100, 1000]. Since it is necessary to determine whether there is a nonlinear cointegration relationship among input variables, the output variable of LS-SVM model is set as an SMM sequence according to the definition of nonlinear cointegration.

5.2.1. Model Parameter Setting. Define key parameters of PSO algorithm: particle number is 20; the particle length is 2, and the parameters to be optimized are $\gamma$ and $\sigma^2$. The search range of parameter $\gamma$ is (0,400000), and search range of $\sigma^2$ is (0,10000). The learning factors $c_1$ and $c_2$ shall be 2. The maximum number of iterations is 5000; the initial value of inertia weight $\omega$ was defined as 0.9, which decreased linearly to 0.4 with the increase of iteration number. The maximum speed $V_{\text{max}}$ is 10. Random initialization of particle swarm in each particle position and velocity, namely $t = 0$, $X_i(0) = (\gamma_i(0), \sigma_i^2(0))$, $V_i(0) = (V_{\gamma_i}(0), V_{\sigma_i^2}(0))$. According to the above settings, the optimal parameters shall be calculated as $\gamma = 34369.670$, $\sigma^2 = 5.912$. The final error with the target value is -0.0251, and the average error is -0.0268. The iteration in the calculation process and the particle distribution in the final iteration are shown in Figure 3 and Figure 4, respectively.

5.2.2. Model Optimization Results. According to the above settings, the optimal parameters shall be calculated as $\gamma = 34369.670$, $\sigma^2 = 5.912$. The final error with the target value is -0.0251, and the average error is -0.0268. The iteration in the calculation process and the particle distribution in the final iteration are shown in Figure 3 and Figure 4, respectively.

5.2.3. Nonlinear Relationships Test. The test samples from 2018 to 2021 were substituted into the LS-SVM model optimized by PSO algorithm, respectively, and the modified R/S statistic $Q_n^*$ was used to perform long memory test for the 4 output sequences. The statistical test results are shown in Table 4.
5.3. Prediction of SMPI Based on NECM Model. Based on the test results of nonlinear cointegration relationship among SMPI, CPI, PPI, and MPI, this article establishes the NECM prediction model of SMPI.

5.3.1. Model Setting. We suppose that \( X_t = (\text{SMPI}_t, \text{CPI}_t, \text{PPI}_t, \text{MPI}_t) \), the nonlinear cointegration function among SMPI, CPI, PPI, and MPI is \( f(X_t) \). The model shall be defined as follows:

\[
\Delta \text{SMPI}_t = \alpha_{10} + \sum_{j=1}^{k_1} \alpha_{1j} \Delta \text{SMPI}_{t-j} + \sum_{j=1}^{k_2} \alpha_{2j} \Delta \text{CPI}_{t-j} \\
+ \sum_{j=1}^{k_3} \alpha_{3j} \Delta \text{PPI}_{t-j} + \sum_{j=1}^{k_4} \alpha_{4j} \Delta \text{MPI}_{t-j} + \alpha_5 f(X_t)
\]

(18)

According to the AIC and SC criteria, the lag order in formula (18) is determined as \( k_1 = 1, k_2 = 3, k_3 = k_4 = 1 \). From the properties of the NECM model, it can be known that the estimation of the nonlinear function \( f(X_t) \) in equation (18) is directly related to the time of the system. As \( t \) is different, the estimation of the nonlinear cointegration function is also different. The change of SMPI at time \( t \), \( \Delta \text{SMPI}_t \), is affected by the estimation result of the system deviation from the equilibrium value \( f(X_{t-1}) \) at time \( t-1 \). Therefore, the lag order \( k_5 \) in the nonlinear function estimation is 1.

5.3.2. Estimation of Model Parameters. To predict SMPI based on the NECM model, it is necessary to estimate the \( f(X_t) \) of the system at different times. Afterward, the estimation result shall be substituted into equation (18) to estimate the parameters of the equation. Therefore, this article also uses the data from 2000 to 2017 as training samples to estimate the nonlinear functional relationship based on the parameters of the LS-SVM model optimized by PSO, applies the output as the estimation result of \( f(X_{2017}) \), and adopts the least squares method. We estimated equation (18) for obtaining \( \Delta \text{SMPI}_{2018} \) in 2018. Afterward, the predicted value of SMPI of 2018 is also obtained by \( \text{SMPI}_t = \text{SMPI}_{t-1} + \Delta \text{SMPI}_t \), with implementing the error analysis for actual value. Along the same lines, the SMPI for 2019–2021 shall be predicted and compared with the actual value using a rolling forecast verification method. In the construction of each \( \Delta \text{SMPI}_t \) prediction model from 2018 to 2021, the LS-SVM model was optimized by the PSO algorithm, and the nonlinear cointegration function among SMPI, CPI, PPI, and MPI is estimated as well. We have the parameters \( \gamma \) and \( \sigma^2 \) in the LS-SVM model. The results are as follows shown in Table 5.

With LS-SVM optimized by PSO in each group, we have sequence results of \( f(X_{t}) \) \( (t = 2017, 2018, 2019, 2020) \) were substituted into Equation (18), and the least-squares estimation of model parameters was carried out by using Eviews software. Taking \( \Delta \text{SMPI}_{2018} \) as an example, the parameter estimation results are shown in Table 6.

5.3.3. Analysis and Comparison of Model Prediction Results. According to the parameter estimation results in Table 6, we use formula (18) to obtain \( \Delta \text{SMPI}_{2018} = -7.469 \). The above results and the ship maintenance cost index in 2017 are substituted into \( \text{SMPI}_t = \text{SMPI}_{t-1} + \Delta \text{SMPI}_t \), and we have \( \text{SMPI}_{2018} = 575.235 \). The error with the actual value of 2018 ship maintenance cost index is 0.26%. According to the same idea, we may have the predicted results of the SMPI from 2019 to 2021 and the error with the actual value. The prediction results and errors of NECM are shown in Table 7.

In order to compare the prediction effect of SMPI based on NECM, this article constructs linear VAR models of SMPI, CPI, PPI, and MPI with samples from 2000 to 2021 to...
predict SMPI and compares the results with NECM. The VAR model is defined as follows:

\[
\text{SMPI}_t = c_{10} + \sum_{j=1}^{k_1} c_{1j} \text{SMPI}_{t-j} + \sum_{j=1}^{k_2} c_{2j} \text{CPI}_{t-j} + \sum_{j=1}^{k_3} c_{3j} \text{PPI}_{t-j} + \sum_{j=1}^{k_4} c_{4j} \text{MPI}_{t-j}.
\] (19)

The lag order of each variable in Formula (19) is set in the same way as that in Formula (18). The samples from 2000 to 2017 are substituted into Equation (19), and the least square method is adopted to estimate model parameters. According to the estimation results of model parameters, we predict the SMPI from 2018 to 2021. The prediction results and errors of VAR model are shown in Table 7.

As can be seen from Table 7, compared with SMPI prediction results based on NECM model, the prediction effect of VAR model is poor and cannot be effectively predicted, which also proves that the linear relationship among SMPI, CPI, PPI, and MPI is not significant.

Next, based on the conclusion that there is a nonlinear relationship among SMPI, CPI, PPI, and MPI, we use the wavelet neural network algorithm proposed in literature [7] to study the nonlinear cointegration relationship, establish the NECM model to predict SMPI, and get the prediction results from 2018 to 2021. The results are shown in Table 7. In the research process, it is found that the calculation process of wavelet neural network algorithm is more complex than LS-SVM. At the same time, due to the small number of samples in this paper (there are about 480 training samples in literature [7]), there is an overfitting phenomenon, but the accuracy of future trend prediction is also weaker than the model proposed in this article.

6. Conclusion

In this article, the small sample nonlinear cointegration test and NECM based on the LS-SVM optimized by PSO were studied. And the logical process of this method was also designed. Then, we carry out empirical research on the SMPI and the several price indexes. Based on the judgment of the type of cointegration test relationship, the test of the nonlinear cointegration test relationship of small samples was realized, the NECM for predicting the SMPI was established as well, which compared with the VAR model of linear vector. The research result showed that the small sample nonlinear cointegration test and modeling method based on the LS-SVM optimized by PSO can describe the nonlinear cointegration test relationship of the small sample system. And the NECM had better performance. The prediction effect can effectively predict small sample nonlinear systems. We also compared the prediction results with the wavelet neural network algorithm, and the results showed that the generalization ability of LS-SVM Optimized by PSO was better, and the prediction accuracy of small samples was higher.

Data Availability

The data presented in this study are available upon reasonable request from the corresponding author.
Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the National Social Science Foundation of China with grant nos. 17BJY028 and 19CGL073.

References

[1] R. F. Engle and C. W. J. Grange, “cointegration and error correction: representation estimation and testing,” Econometrica, vol. 55, no. 3, pp. 251–276, 1987.
[2] M. Rahman and A. Islam, “Some dynamic macroeconomic perspectives for India’s economic growth: applications of linear ARDL bounds testing for cointegration and VECM,” Journal of Financial Economic Policy, vol. 12, no. 4, pp. 641–658, 2020.
[3] C. Mouelhi, “The relationship between cat bond market and other financial asset markets: evidence from cointegration tests,” European Journal of Business and Management Research, vol. 6, no. 2, pp. 78–85, 2021.
[4] M. Sulochana and S. Rajkumar, “An empirical analysis of cointegration test among asian stock markets,” Test Engineering and Management, vol. 83, pp. 9619–9625, 2021.
[5] C. W. J. Granger, “Modelling nonlinear relationships between extended memory variables,” Econometrica, vol. 63, no. 2, pp. 265–279, 1995.
[6] X. B. Zhang, Q. H. Sun, and S. Y. Zhang, “Research on nonlinear cointegration test and its testing method,” Journal of system engineering, vol. 14, no. 1, pp. 57–68, 1999.
[7] X. Sun, G. Wang, and Y. Fan, “Adaptive trajectory tracking control of vector propulsion unmanned surface vehicle with disturbances and input saturation,” Nonlinear Dynamics, vol. 106, no. 3, pp. 2277–2291, 2021.
[8] L. Sun, S. Chen, J. Xu, and T. Yun, “Improved Monarch Butterfly Optimization Algorithm Based on Opposition-Based Learning and Random Local Perturbation,” Complexity, Article ID 4182148, vol. 2019, 20 pages, 2019.
[9] V. K. Kamboj, C. L. Kumari, S. K. Bath et al., “A cost-effective solution for non-convex economic load dispatch problems in power systems using slime mould algorithm,” Sustainability, vol. 14, no. 5, p. 2586, 2022.
[10] A. M. Hussein, M. Abd Elaziz, M. S. A. Wahed, and M. Sillanpää, “A new approach to predict the missing values of algae during water quality monitoring programs based on a hybrid moth search algorithm and the random vector function link network,” Journal of Hydrology, vol. 575, pp. 852–863, 2019.
[11] I. Davut, E. Serdar, E. Erdal, and K. Murat, “A novel modified opposition-based hunger games search algorithm to design fractional order proportional-integral-derivative controller for magnetic ball suspension system,” Advanced Control for Applications: Engineering and Industrial Systems, vol. 4, no. 1, 2022.
[12] A. Iman, H. A. Asghar, H. Gandomi Amir, C. Xuefeng, and C. Huiling, “RUN beyond the metaphor: an efficient optimization algorithm based on Runge Kutta method,” Expert Systems with Applications, vol. 181, Article ID 115079, 2021.
[13] J. Tu, H. Chen, M. Wang, and A. H. Gandomi, “The Colony predation algorithm,” Journal of Bionics Engineering, vol. 18, no. 3, pp. 674–710, 2021.
[14] S. Song, P. Wang, A. A. Headari, X. Zhao, and H. Chen, “Adaptive Harris hawks optimization with persistent trigonometric differences for photovoltaic model parameter extraction,” Engineering Applications of Artificial Intelligence, vol. 109, Article ID 104608, 2022.
[15] Y. Feng, S. Deb, G. G. Wang, and A. H. Alavi, “Monarch butterfly optimization: a comprehensive review,” Expert Systems with Applications, vol. 168, Article ID 114418, 2021.
[16] J. Alfadhl, A. Jaragh, M. G. Alfajalaki, and I. Ahmad, “FP-SMA: an adaptive, fluctuant population strategy for slime mould algorithm,” Neural Computing & Applications, vol. 34, no. 13, Article ID 11175, 2022.
[17] A. Fathy, M. A. Elaziz, E. T. Sayed, A. G. Olabi, and R. Hegazy, “Optimal parameter identification of triple-junction photovoltaic panel based on enhanced moth search algorithm,” Energy, vol. 188, no. 1, pp. 1–14, Article ID 116025, 2019.
[18] B. Shi, H. Ye, L. Zheng et al., “Evolutionary warning system for COVID-19 severity: Colony predation algorithm enhanced extreme learning machine,” Computers in Biology and Medicine, vol. 136, Article ID 104698, 2021.
[19] J. Liu, X. Liu, Y. Wu, Z. Yang, and J. Xu, “Dynamic multi-swarm differential learning Harris hawks optimizer and its application to optimal dispatch problem of cascade hydro-power stations,” Knowledge-Based Systems, vol. 242, Article ID 108281, 2022.
[20] S. R. Fahim, H. M. Hasanien, R. A. Turky et al., “Parameter identification of proton exchange membrane Fuel cell based on hunger games search algorithm,” Energies, vol. 14, no. 16, pp. 1–21, 2021.
[21] M. Yang, B. Ai, R. He et al., “Machine-learning-based fast angle-of-arrival recognition for vehicular communications,” IEEE Transactions on Vehicular Technology, vol. 70, no. 2, p. 1, 2021.
[22] Y. Li, H. Darong, and Q. Zixia, “A Classification Algorithm of Fault Modes-Integrated LSSVM and PSO with Parameters’ Optimization of VMD,” Mathematical Problems in Engineering, vol. 2021, Article ID 6627367, 12 pages, 2021.
[23] D. Lee and P. Schmidt, “On the power of the KPSS test of stationarity against fractionally-integrated alternatives,” Journal of Econometrics, vol. 73, no. 1, pp. 285–302, 1996.
[24] P. M. Robinson, “Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression,” Journal of Econometrics, vol. 47, no. 1, pp. 67–84, 1991.
[25] M. Ahmadi and Z. Chen, “Machine learning-based models for predicting permeability impairment due to scale deposition,” Journal of Petroleum Exploration and Production Technology, vol. 10, no. 7, pp. 2873–2884, 2020.
[26] M. A. Ahmadi, “Toward reliable model for prediction Drilling Fluid Density at wellbore conditions: a LSSVM model,” Neurocomputing, vol. 211, no. 26, pp. 143–149, 2016.
[27] M. A. Ahmadi and M. Pournik, “A predictive model of chemical flooding for enhanced oil recovery purposes: application of least square Support vector machine,” Petroleum, vol. 2, no. 2, pp. 177–182, 2016.
[28] M. A. Ahmadi and Z. Chen, “Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs,” Petroleum, vol. 5, no. 3, pp. 271–284, 2019.
[29] M. A. Ahmadi and A. Bahadori, “A LSSVM approach for determining well placement and conning phenomena in horizontal wells,” Fuel, vol. 153, no. 1, pp. 276–283, 2015.
[30] N. I. U. Wening, H. A. Yinaer, and C. Nan, “Support vector machine based machine learning method for GS 8QAM constellation classification in seamless integrated fiber and...
visible light communication system,” *Science China (Information Sciences)*, vol. 63, no. 10, pp. 182–193, 2020.

[31] B. P. Vrigazova, “Detection of malignant and benign breast cancer using the ANOVA-BOOTSTRAP-SVM,” *Journal of Data and Information Science*, vol. 5, no. 2, pp. 62–75, 2020.

[32] Y.-B. Shen, J. Song, and W. Zhi-chao, “SVM feature selection and parameter optimization based on improved fireworks algorithm,” *Microelectronics and Computers*, vol. 35, no. 1, pp. 21–25, 2018.

[33] Y. L. Gu, Y. Zhang, X. P. Rui, W. Q. Lu, M. Li, and S. Wang, “Short-term traffic flow prediction based on LSSVM optimized by immune algorithm,” *Journal of Jilin University (Engineering and Technology Edition)*, vol. 49, no. 06, pp. 1852–1857, 2019.

[34] C. Tian, Z. Xing, X. Pan, and H. A. Wang, “A GA-LSSVM approach for predicting and controlling in screw chiller,” *Proceedings - Institution of Mechanical Engineers Part A Journal of Power and Energy*, vol. 235, no. 7, 2021.

[35] E. S. El-Kenawy and M. Eid, “Hybrid gray wolf and panicle swarm optimization for feature selection,” *International Journal of Innovative Computing Information and Control*, vol. 16, no. 3, pp. 831–844, 2020.

[36] B. B. Mandelbrot and M. Taqqu, “Robust R/S analysis of long run serial correlation,” *Bulletin of International Statistical Institute*, vol. 48, no. 3, pp. 59–104, 1979.

[37] A. Lo, “Long-term memory in stock market prices,” *Econometrica*, vol. 59, no. 5, pp. 1279–1313, 1991.