sustainability

MDPI

Article

Forecasting Educational Inequality in China for Sustainable Development: A Hybrid Framework of GM(1,1) and CS-SVR

Zhe Gao [1], Tianxiang Shi [2] and Lihao Shang [1,*]

[1] College of Education, Zhejiang University, Hangzhou 310058, China; gaozhe@zju.edu.cn
[2] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China; 22012256@zju.edu.cn
* Correspondence: shlihao@zju.edu.cn

Abstract

Educational equality is essential for achieving social justice and sustainable development. Accurately predicting the trend of educational inequality is important for improving education systems and ensuring equitable resource allocation. In this paper, the Educational Gini (E-Gini) index is calculated based on the population aged 6 and above in China from 2002 to 2024, quantifying educational inequality. To forecast the future trend in the E-Gini index, a hybrid prediction framework based on the grey prediction model (GM(1,1)) and Cuckoo search-support vector regression (CS-SVR) model is proposed. This framework incorporates three influencing factors, including government budget spending on education, per capita consumption expenditure on education, and the Consumer Price Index (CPI) for education. The results show that the E-Gini of China generally declines from 2002 to 2024 with fluctuations. The proposed approach predicts the E-Gini value of 2024 as 0.220130, while the actual value is 0.2206, corresponding to an absolute error of 0.000470 and a relative error of 0.213%. In the benchmark comparison, the proposed model outperforms the linear trend model, the univariate GM(1,1), the naive persistence model, ARIMA, and the standard SVR model. The comparative analysis demonstrates that the proposed framework effectively captures the inherent patterns of educational inequality and reveals its trends. The proposed framework serves as a valuable tool for forecasting trends in educational inequality and informing policy decisions.

Keywords: cuckoo algorithm; support vector regression; GM(1,1); educational inequality; E-Gini index

Check for updates

Academic Editor: Adela García-Aracil

Received: 20 March 2026
Revised: 23 April 2026
Accepted: 23 April 2026
Published: 25 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Educational inequality has become a critical issue in the pursuit of sustainable development. It not only constrains human capital formation but also undermines social well-being and long-term social equity. The 2030 Agenda for Sustainable Development clearly states that "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" is one of the Sustainable Development Goals for all countries. Other studies indicate that educational inequality has a significant influence on the overall well-being of citizens, economic productivity [1], the accumulation of human capital [2], and even intergenerational achievement. Therefore, research on educational inequality has both theoretical and practical significance for understanding social structural transformations and guiding the development of educational policies.

Scholars typically define educational inequality from the perspectives of inequality in educational opportunities [3–5] and inequality of educational outcomes (or educational

attainment) [6–8]. To measure educational inequality, the Education Gini (E-Gini) index has recently been proposed. This index has been widely used, and its research value and rationality have been validated [9–12].

Predicting the changing trends of educational inequality serves as an important reference for policy formulation, education reform, and other related work. From a research perspective, most existing studies have attempted to capture the evolution of educational inequality from two primary approaches. One approach focuses on the specific details of educational inequality, such as financial inequalities [13], racial educational disparities [14,15], and inequalities in educational effectiveness [16]. The other approach is to examine the factors influencing macro-trends in education [17], such as policy fluctuations [18], economic uncertainty [19], quality of governance [20], and digital technology [21]. Since educational inequality is affected by various factors, selecting the relevant and available data is necessary for predicting its trends. Therefore, national budget spending on education [22,23] and educational inputs of households [24,25] are chosen.

The univariate time-series forecasting models are often inappropriate to predict educational inequality, which is influenced by a multitude of factors [26]. The multivariate time-series method can take more influencing factors into account in the prediction model and improve the practical significance of the results. However, the related data released by the government usually lags [27], leading multivariate time series forecasts to face a lack of data on influencing factors. In recent years, the proposal of new forecasting frameworks in the economic and engineering fields has provided a solution to this dilemma [28,29]. This hybrid method makes the prediction results more targeted and is expected to solve the problem of data lag impacting the prediction target.

There are still gaps in the use of the predictive approach in education [26]. Most existing studies take overall educational indicators (including enrollment, educational investment, and overall development levels) as prediction targets, while there are few prediction studies that directly use the E-Gini index as the target variable [30]. Moreover, though educational inequality is affected by multiple factors, current approaches usually neglect the reality factors [17]. The single-variable prediction method based on numerical patterns cannot provide explanations for the prediction results. Furthermore, in cases where the annual sample sequence is short and official statistics are lagging behind, traditional multivariate methods are no longer appropriate either [27]. Since the national educational statistics are usually characterized by small data volumes and high volatility, Support Vector Regression (SVR) is a suitable solution for predicting the development trends of the E-Gini index. As a widely used regression prediction method, the SVR model maps the training sample to a higher dimensional feature space for better fitting [31]. SVR handles the complex relationships between input and output variables and is more suitable for solving nonlinear, small-sample, or high-dimensional problems than machine learning algorithms such as artificial neural networks [32]. However, the penalty parameter C and the parameter of the kernel function g, which have significant effects on the model performance, are determined differently [32,33]. The meta-heuristic optimization techniques have been proposed to search for the optimal parameters to maximize the model's benefits [34,35]. The proposed framework uses the cuckoo search (CS) algorithm [36] to determine the optimal model parameters for the SVR model. This approach is termed the CS-SVR model in this paper. This study employs the GM(1,1) model to forecast the short-term variation of explanatory variables under small-sample conditions. Furthermore, CS-SVR is applied to capture the nonlinear relationship between the predicted variables and the E-Gini index. The study proposes a novel approach to predicting educational inequality. The study is structured around three research questions:

(1) How does educational inequality measured by the E-Gini index in China evolve over the period 2002–2024?
(2) Can the proposed framework achieve accurate predictions of the E-Gini of China under annual small-sample conditions after incorporating reality-related factors?
(3) Can the proposed framework provide a more informative basis for monitoring future educational inequality and supporting the sustainable improvement of educational equity?

A new hybrid framework is proposed in this paper. This framework combines the GM(1,1) model and the CS-SVR model to predict the trend of the E-Gini index by incorporating the national education input, funds channeled into education using the national budget, and the Consumer Price Index. In addition, this paper discusses the applicability of the GM(1,1) model to the datasets mentioned before and validates the accuracy and effectiveness of the proposed framework. The main contributions of this paper can be categorized as:

(1) We provide a national time-series description of educational inequality in China based on the E-Gini index for 2002–2024.
(2) We develop and validate a two-stage forecasting framework for short-term E-Gini prediction under annual small-sample conditions.
(3) We evaluate whether incorporating government budget spending on education, per capita consumption expenditure on education, and CPI for education can improve the practical value of educational inequality forecasting for sustainability-oriented monitoring.

The remainder of this paper is structured as follows. Section 2 reviews related prior work. Section 3 presents the data and methods. Section 4 reports and discusses the empirical results. Section 5 concludes the paper, addresses the limitations of the present study, and provides suggestions for future research.

2. Literature Review

Predicting educational trends has attracted the attention of scholars with the development of computer technology. For example, Li et al. [26] construct a weakening buffer operator-based GM(1,1) model to predict the gross enrollment rate of higher education in Kazakhstan. Li [37] uses Chinese university enrollment data from 2010 to 2020 and applies polynomial regression and Holt's exponential smoothing to forecast enrollment trends, chain growth rates, and changes in undergraduate enrollment proportions across regions. Most of the presented work focuses on other characteristics of education. For instance, Liang et al. [38] use the L-M algorithm to predict the scale of public investment in higher education and analyze the configuration structure among different types of higher education institutions. Friedman et al. [39] model the within-country distribution of years of schooling and use their model to predict the achievement of Sustainable Development Goals (SDG). They suggest that the universalization of primary education could be initially achieved by 2030, but significant challenges remain for secondary and tertiary education. Maliti [40] uses six rounds of methodological stable Demographic and Health Surveys (DHS) to demonstrate that horizontal inequalities in education and wealth have been declining in Tanzania.

Trends in educational inequality are influenced by multiple factors. This complexity calls for meaningful forecasting using multivariate time-series methods. There is a statistically significant correlation between the proportion of national and household expenditures on education and the E-Gini coefficient [41–43]. Xu et al. [17] use a Nonlinear Autoregressive Distributed Lag (NARDL) model to reveal the existence of asymptotic effects on the Gini index and E-Gini index in China over the period 1975–2020. Therefore, it is essential to incorporate these economic factors into the forecasting framework.

The hybrid forecasting framework helps to mitigate the issue that the influencing factors in the future are missing. For example, Zhu et al. [29] employ a mixed model to predict the credit risk of Small and Medium-Sized Enterprises (SMEs) in China. Similarly, Liang et al. [28] predict the development of offshore wind energy in Europe using a two-stage forecasting framework and demonstrate its effectiveness.

The existing literature provides useful but incomplete guidance for the present study. Educational forecasting studies have predicted indicators such as enrollment, educational investment, and overall development levels, but they mainly focus on educational development rather than the distributional dimension of education, especially E-Gini-based inequality. Studies on educational inequality provide important evidence on measurement and associated factors, but they rarely develop forecasting strategies for the E-Gini index. Hybrid forecasting studies in other fields demonstrate that model combination can improve predictions under complex conditions, but these approaches have not been specifically adapted to the problem of forecasting educational inequality.

The present study addresses this gap by adapting a hybrid forecasting strategy to the specific requirements of E-Gini prediction. The proposed framework links the forecasting of macro-level predictors to the prediction of educational inequality, thereby moving beyond approaches that rely only on the internal pattern of the E-Gini series. It also addresses the practical problem that future values of key explanatory variables are not directly available at the time of forecasting. In this sense, the study contributes a forecasting design tailored to educational inequality under realistic data constraints.

3. Data and Methods

This section presents the proposed framework. Educational inequality is influenced by many factors, making it inadequate to rely on a single data sequence to predict the development of the E-Gini Index. Ignoring these influencing factors will lead the forecast results to lose their practical significance. To address these issues, this paper proposes a hybrid framework that combines the GM(1,1) model and the CS-SVR model, integrating the impact of the above factors on the trend of educational inequality. It should be noted that the proposed framework can be extended to consider other factors that affect educational inequality. Specifically, the proposed framework first predicts the trends of three key indicators (i.e., government budget spending on education, per capita consumption expenditure on education, and CPI for education) using the GM(1,1) model.

The input variables are chosen since they can capture three macro-level dimensions associated with educational inequality at the national level. Government budget spending on education reflects the public financing capacity of the education system and the country's potential compensatory ability. Prior research shows that educational spending is related to educational inequality, although the magnitude and direction of this relationship vary across contexts [41,44].

Per capita consumption expenditure on education can capture families' private investment in children's education and the economic burden they bear in acquiring educational opportunities [4]. In research of China, existing studies further suggest that household educational expenditure is closely associated with inequality, opportunity structures, and competition in shadow education [42]. Moreover, the education CPI is introduced as a macro-level proxy for changes in education-related prices in this paper. Though it does not directly measure educational inequality, it reflects shifts in the cost environment facing both households and the state in education-related decision-making [45]. Increases in education-related costs may affect educational affordability and widen disparities in effective access to educational resources. Actually, these three indicators do not fully capture all determinants of educational inequality. However, the proposed framework has the

potential to incorporate other factors in future research. The limitations of the present study are discussed separately in the concluding section.

Simultaneously, the relationship between these three independent indicators and the E-Gini index is modeled using the CS-SVR model. Finally, based on these three validated predictions, the CS-SVR model forecasts the trend of educational inequality in the next year. By combining and optimizing multiple methods, this framework enhances the accuracy and stability of predictions, providing more meaningful results [28,46–48]. The flow chart of the proposed framework is shown in Figure 1.

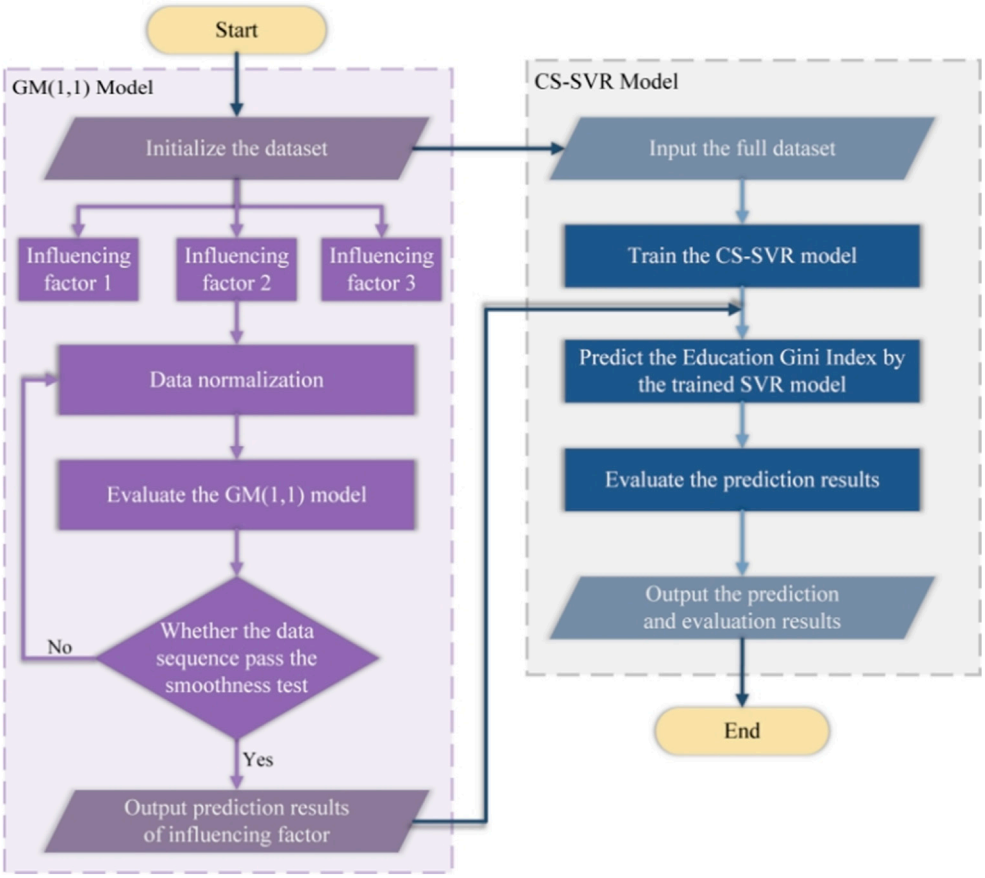


Figure 1. The flow chart of the proposed framework.

3.1. The Modelling Algorithm of GM(1,1) Forecasting Model

Grey theoretical models typically use the accumulated generating operation (AGO) technique to minimize the randomness of the original data. This method extracts potential information from the data by solving differential equations [26].

First, an original data sequence is assumed as

$$X_0 = \left\{ x_0(1), \quad x_0(2), \quad \ldots, \quad x_0(n) \right\} \tag{1}$$

An operator is derived through the first-order accumulated generating operation (1-AGO) to smooth the original data and reduce fluctuations. This allows the model to better capture potential trends.

$$X_1 = \left\{ x_1(1), \quad x_1(2), \quad \ldots, \quad x_1(n) \right\} \tag{2}$$

with

$$x_1(k) = \sum_{i=1}^{k} x_0(i),\ k = 1, 2, \ldots, n \tag{3}$$

The accumulation sequence X_1 follows the first-order linear differential equation

$$\frac{dx_1(k)}{dt} + ax_1(k) = b \tag{4}$$

where a is the development coefficient, which indicates the trend of decay or growth of the system and b is the grey input. Equation (3) is often known as the whitening GM(1,1) model.

The discrete Equation (3) yields an approximation of $x_1(k)$, which is defined as

$$z_1(k) = \alpha x_1(k) + (1 - \alpha)x_1(k-1) \cong x_1(k) \tag{5}$$

When α in Equation (5) is set to 0.5, the GM(1,1) model simplifies to

$$x_0(k) + az_1(k) = b \tag{6}$$

Estimating the parameters a and b is an important step, and they are commonly determined using the least-squares method applied to the GM(1,1) model

$$\hat{\theta} = \begin{bmatrix} a \\ b \end{bmatrix} = \left[B^T B\right]^{-1} B^T Y \tag{7}$$

with

$$B = \begin{bmatrix} -z_1(2) & 1 \\ -z_1(3) & 1 \\ \vdots & \vdots \\ -z_1(n) & 1 \end{bmatrix},\ Y = \begin{bmatrix} x_0(2) \\ x_0(3) \\ \vdots \\ x_0(n) \end{bmatrix} \tag{8}$$

After obtaining the parameters a and b, substituting $\hat{\theta} = [a, b]^T$ into Equation (4) and based on the initial conditions $x_1(1) = x_0(1)$, one can finally obtain a special solution of Equation (4)

$$\hat{x}_1(t) = \left(x_0(1) - \frac{b}{a}\right)e^{-a(t-1)} + \frac{b}{a},\ \ t \in [1, +\infty) \tag{9}$$

Here $\hat{x}_1(t)$ represents the grey fitted value of $x_1(t)$.

The discrete form of Equation (9) is defined as the grey forecasting model, which is expressed as

$$\hat{x}_1(k+1) = \left(x_0(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a},\ \ k = 1, 2, \ldots, n-1 \tag{10}$$

Finally, the prediction of the original data is obtained by the Inverse Accumulated Generating Operation (IAGO)

$$\hat{x}_0(k+1) = \hat{x}_1(k+1) - \hat{x}_1(k),\ \ k = 1, 2, 3, \ldots, n \tag{11}$$

The above derivation process reveals that the GM(1,1) model does not rely on large datasets, but is instead more suitable for short-term forecasting with limited data volumes [49]. In addition, the currently publicly available data about education does not meet the requirements for large datasets needed to train traditional neural network models. The characteristics of the GM(1,1) model align precisely with our data. Therefore, the trends of the underlying variables in this paper are predicted using the GM(1,1) model.

3.2. Cuckoo Search-Support Vector Regression (CS-SVR) Model

The SVR model demonstrates good performance in learning problems with small data samples and can reduce the uncertainty of assessment results. The education-related datasets collected in this paper are annual, which cannot meet the training requirements of traditional neural networks. Therefore, the SVR model is used to address this challenge. However, SVR models are usually inefficient due to the subjective nature of selecting key parameters in practice [50].

To further improve the prediction accuracy, the Cuckoo Search (CS) bionic algorithm is introduced to optimize the parameters of the SVR model, which is described in the following section. The flowchart of the CS-SVR model is presented in Figure 2. This method has received much attention in engineering [35] and economic fields [51,52]. Compared with commonly used models such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA), the CS algorithm is more efficient [32]. In this paper, CS-SVR is used to predict the trend of educational inequality.

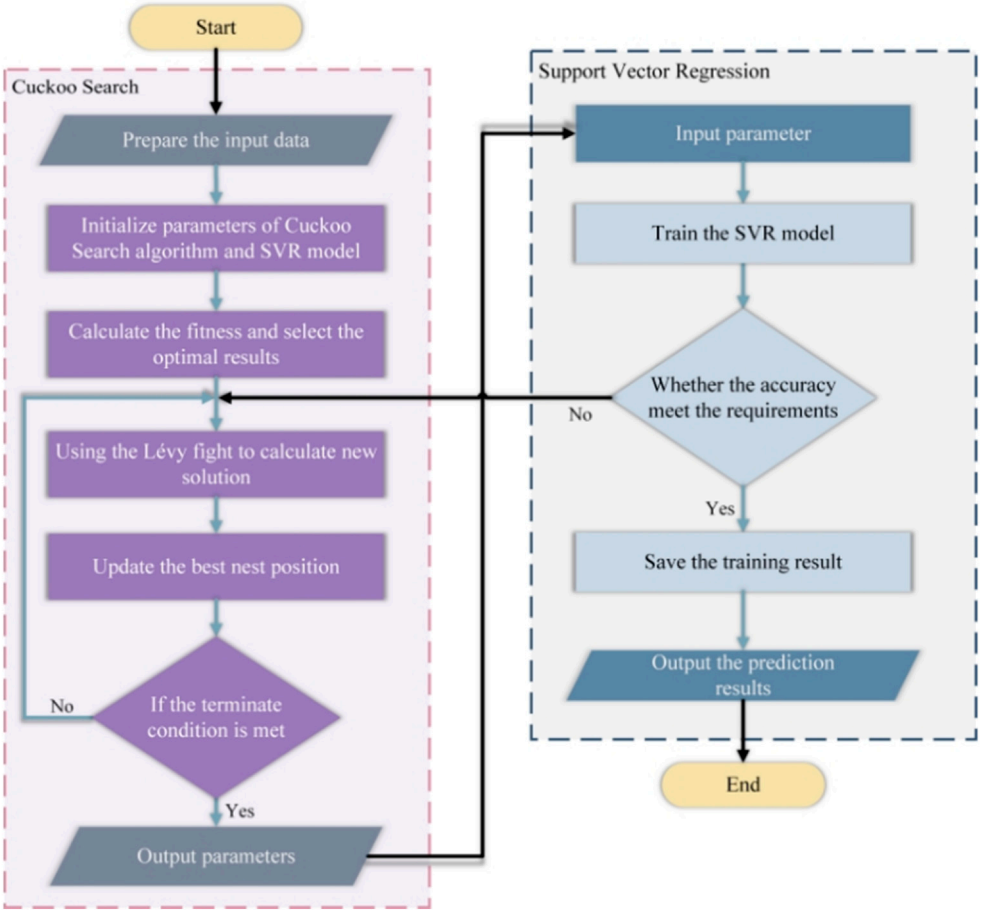


Figure 2. The flow chart of CS-SVR.

3.2.1. Cuckoo Search (CS) Bionic Algorithm

The cuckoo bionic algorithm is a metaheuristic optimization algorithm that mimics the unique brood parasitism behavior of cuckoos [53]. In nature, cuckoos exhibit brood parasitism by laying their eggs in the nests of other host birds. The cuckoo chicks increase their survival chances by hatching earlier than the host's offspring and eliminating competition for resources [54,55]. In this bionic algorithm, each nest represents a candidate solution, and a cuckoo's egg represents a new, potentially better solution.

During the optimization process, the algorithm employs Lévy flights [54] to explore the search space and generate new solutions, which can be mathematically represented as

$$y_{ij}(t+1) = y_{ij}(t) + \alpha \times Levy(\lambda) \tag{12}$$

where $y_{ij}(t)$ represents the current position of the nest, $y_{ij}(t+1)$ represents the new position of the nest, α is the step size, and λ is the parameter of the Lévy flight. The process of searching for the optimal solution can be observed in Figure 2.

The Gaussian Radial Basis Function (RBF) kernel for SVR applications provides good generalization on local scales [32]. Moreover, the choice of kernel function significantly impacts the performance of SVR models. Therefore, using the CS algorithm to select the optimal parameters is essential for efficient parameter tuning.

3.2.2. Support Vector Regression (SVR) Model

The SVR model is a machine learning method that maps input vectors into a high-dimensional feature space using nonlinear kernel functions. It constructs a regression model based on the principle of structural risk minimization to predict the behavior of variables [48]. The SVR has demonstrated excellent performance in regression problems, especially under small-sample conditions [56], which can be expressed mathematically as

$$f(x) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \tag{13}$$

where **x** represents the input vector, **w** d represents the weight vector, b represents the deviation, and $\phi(x)$ represents the mapping function that maps the samples in the input control to a higher dimensional space.

Using the structural risk minimization principle, the original problem is transformed into

$$\min\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right] \tag{14}$$

where $\|w\|^2$ describes the complexity of the function $f(x)$. C is called the penalty coefficient that controls the model's tolerance to errors. ξ_i and ξ_i^* are slack variables used to measure the deviation outside the ε-insensitive zone. The constraints that must be satisfied are

$$\begin{gathered} y_i - w_i - b \le \varepsilon + \xi_i \\ w_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{gathered} \tag{15}$$

where y_i is the true value corresponding to sample x_i.

To obtain the support vector nonlinear regression function, it is necessary to introduce the concept of Lagrange multipliers to construct the Lagrange function.

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K(x, x_i) + b \tag{16}$$

where α_i^* and α_i represent the Lagrange multipliers. $K(x, x_i)$ represents the kernel function, which is a Gaussian radial basis function in the model. It can be mathematically represented as

$$K(x, x_i) = \exp\left\{-\frac{\|x - x_i\|^2}{g^2}\right\} \tag{17}$$

where g is the key parameter of the kernel function.

3.3. Implementation Details

To enhance methodological transparency, this study adopts a validation design that clearly distinguishes the training period from the out-of-sample evaluation process. The E-Gini series and the three explanatory variables are aligned over the period 2002–2024. The observations from 2002 to 2023 are used for model training, while the 2024 observation is reserved as the final out-of-sample test point. This design can avoid the reported 2024 forecasting results from being an in-sample fit. For the CS-SVR component, the cuckoo search algorithm is employed to optimize two hyperparameters. The search process is configured with 25 nests and 200 iterations. For each candidate parameter combination, predictive performance is evaluated within the training sample using five-fold cross-validation. Moreover, the cross-validation loss is taken as the fitness value. After hyperparameter tuning is completed, the final SVR model is re-estimated using the full training sample from 2002 to 2023, and the resulting model is then used to forecast the E-Gini index of 2024. Forecasting performance is evaluated using MSE, RMSE, and MAE.

To reduce the risk of overfitting, the 2024 observation is excluded from both model training and hyperparameter tuning. It should be noted that the hyperparameters are determined through cross-validation within the training sample, rather than being selected directly on the basis of full-sample fitting results.

4. Results and Discussion

4.1. Predicting Results

China Statistical Yearbook and China Educational Finance Statistical Yearbook are both authoritative official statistical sources. However, their data are subject to a certain publication lag, and some indicators may be adjusted in subsequent releases following statistical verification or population census revisions. In using these data, this study further checks and adjusts them as necessary. To calculate the E-Gini index, educational attainment is categorized into seven levels: no schooling, elementary school graduates, junior high school graduates, senior high school graduates, college degrees, bachelor's degrees, and postgraduate degrees. It is important to note that scholars hold different views on the minimum age for statistical purposes in the group of individuals without formal schooling. Some scholars, such as [11,57], set the minimum age at 15. In this paper, we set the minimum age at 6 years old because the compulsory education in China begins at the primary school level [58,59].

Therefore, the population aged 6 and above who have not yet received formal education is meaningful for identifying inequality in educational opportunity, because it may reflect delayed enrollment, non-enrollment, or disadvantage at the starting point of compulsory education. Compared with some studies that use the population aged 15 and above, the specification adopted in this study is more sensitive to inequalities at the school-entry stage and during compulsory education, and more suitable for the Chinese situation.

It should be noted that E-Gini values calculated using different age thresholds differ in meaning. Therefore, the results of this study are more suitable for longitudinal comparison within China under a consistent measurement framework, and should not be directly compared in absolute terms with existing studies that use the population aged 15 and above. Moreover, the E-Gini is unable to capture differences in educational quality, school resource allocation, or learning outcomes. It should therefore be understood as a conditional description of educational equity.

The data used in this paper, including government budget spending on education, per capita consumption expenditure on education, CPI for education, and total population data, are from the China Statistical Yearbook.

In this case, the E-Gini index serves as the output of the model, while other data series serve as the input of the model. It should be noted here that the E-Gini in this paper is used to quantify educational inequality in China. The relevant data are shown in Table 1. Additionally, the initial input series of the GM(1,1) model undergoes smoothing. The smoothed results are validated to ensure prediction accuracy.

Table 1. Results of the E-Gini index and relevant data of China, 2002–2024.

	Education Gini	Government Budget Spending on Education (Ten Trillion Yuan)	Per Capita Consumption Expenditure on Education (Yuan)	CPI for Education (2002 = 100)
2002	0.2436	0.0349	487	100
2003	0.2397	0.0385	527	104.3
2004	0.2377	0.0447	586	107.84
2005	0.2410	0.0516	657	112.89
2006	0.2357	0.0635	718	113.34
2007	0.2308	0.0828	787	113.35
2008	0.2290	0.1045	814	113.46
2009	0.2251	0.1223	896	115.27
2010	0.2079	0.1467	1000	116.89
2011	0.2123	0.1859	1136	118.41
2012	0.2121	0.2315	1262	120.42
2013	0.2113	0.2449	1398	123.67
2014	0.2165	0.2642	1536	126.64
2015	0.2222	0.2922	1723	130.058
2016	0.2235	0.3140	1915	133.18
2017	0.2219	0.3421	2086	137.17
2018	0.2258	0.3700	2226	141.15
2019	0.2260	0.4005	2513	145.53
2020	0.2205	0.4291	2032	148.73
2021	0.2219	0.4584	2599	151.85
2022	0.2170	0.4847	2469	155.042
2023	0.2192	0.5044	2902	157.21
2024	0.2206	0.5416	3189	159.57

The fitness curve of CS-SVR is shown in Figure 3. The optimal penalty factor, $C = 20.8975$, is obtained under the best fitness. The optimal parameter of the Gaussian kernel function is set to 4.4785.

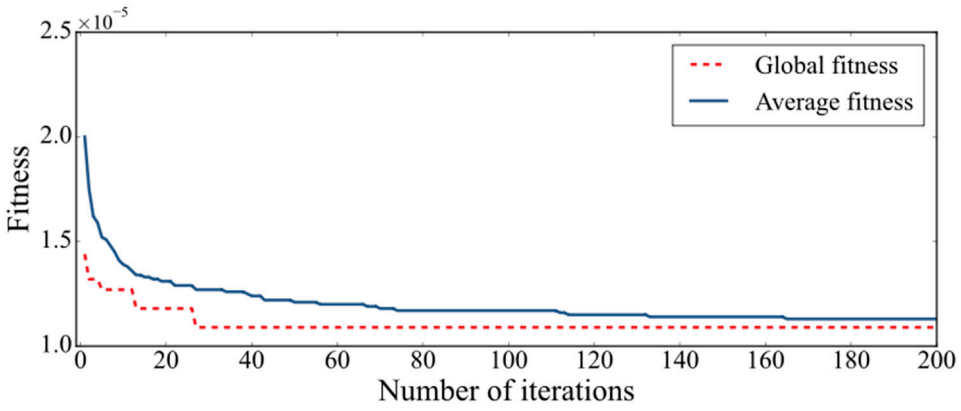


Figure 3. The fitness curve of CS-SVR (Termination of iteration = 200, Number of nests = 25).

Figure 4 presents trends in government budget spending on education and per capita consumption expenditure on education. To show the relationship between two variables, their values are normalized to the same order of magnitude. It can be observed that both variables follow a similar trend. The growth rate of government budget spending on education is significantly lower than the growth rate of per capita consumption expenditure on education. It is worth noting that there is a brief surge in government investment in education around 2010. Figure 5 illustrates the change in CPI for education during this period. Since 2008, the CPI for education has shown a steady upward trend, indicating an increase in education-related prices. This pattern is consistent with prior evidence that household educational expenditure is associated with income inequality and education competition [42].

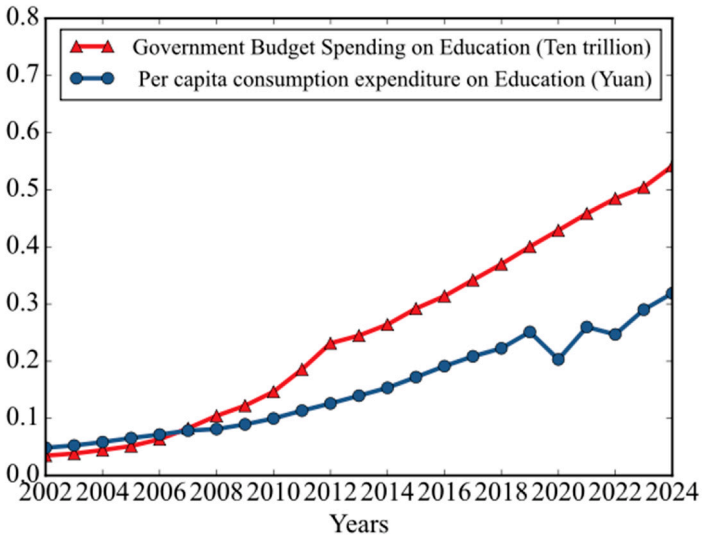


Figure 4. Trends in government budget spending on education and per capita consumption expenditure on education.

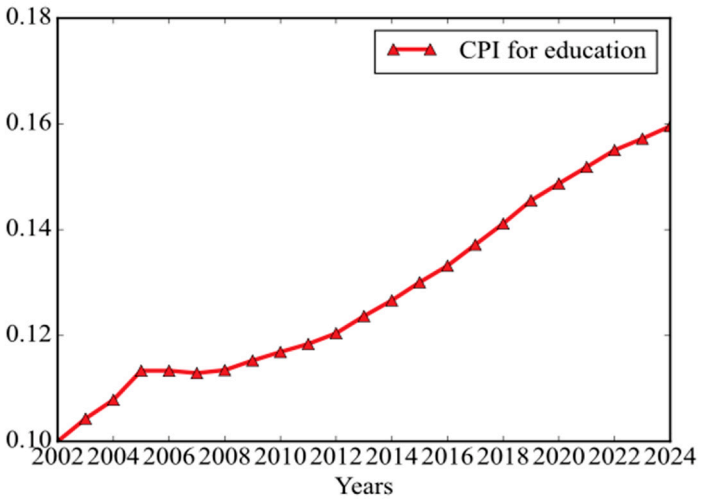


Figure 5. Trends in China's CPI for education, 2002–2024.

To illustrate the superiority of the proposed methods, the same dataset is also simulated using the SVR model. Figure 6 presents the predictions of both the SVR model and the CS-SVR model compared to the annual E-Gini index. Figure 7 illustrates a comparative plot of the relative errors.

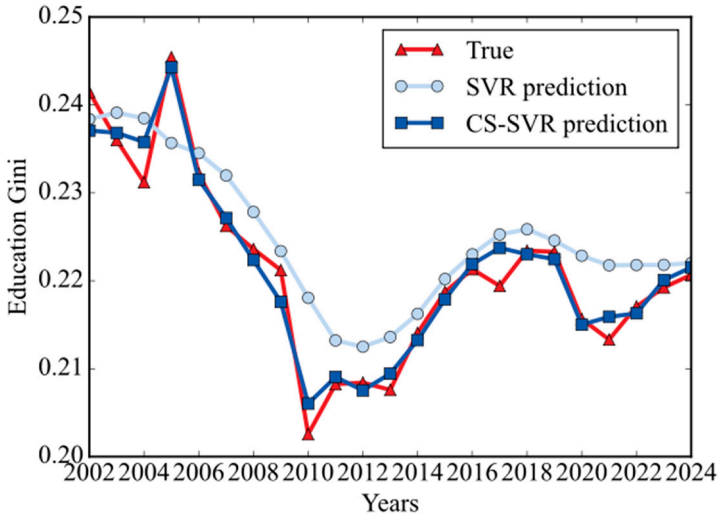


Figure 6. Results of actual and forecast E-Gini index, 2002–2024.

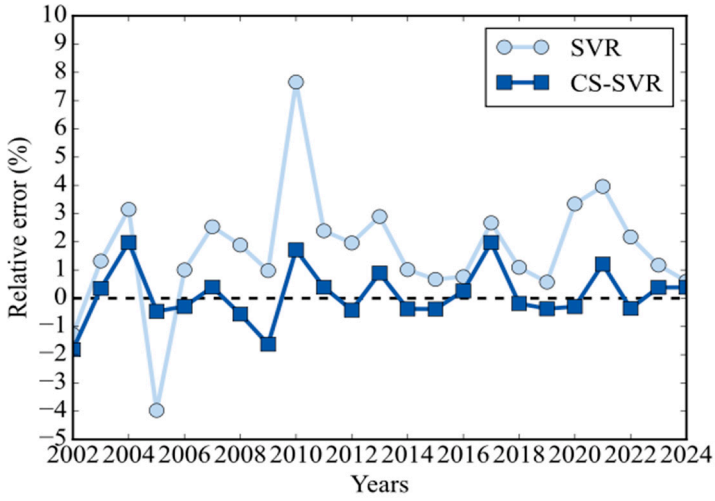


Figure 7. Relative error of the E-Gini index obtained from SVR and CS-SVR.

It can be observed from Figures 6 and 7 that while the prediction results obtained from the SVR model are smoother, they deviate more significantly from the actual data, resulting in higher relative errors. In certain anomalous years, such as 2010, the prediction results show significant deviations from the actual data. This may be caused by changes and biases in the way government statistics are produced. In contrast, the CS-SVR model captures the dataset's trends more effectively, with relative errors generally below 0.5 percent, except for 2010. Under the reported validation design, the CS-SVR model shows desirable predictive performance after incorporating multiple explanatory variables.

To provide a more rigorous evaluation, the proposed framework is compared with several benchmark models, including the naive persistence, the linear trend, ARIMA, the univariate GM(1,1), and the standard SVR.

The univariate benchmarks (including Naive Persistence, Linear Trend, ARIMA, and Univariate GM(1,1)) rely exclusively on the historical E-Gini series and generate the forecasting results for 2024 using 2002–2023 observations. The multivariate benchmark is Standard SVR, which follows the same experimental design as the proposed framework. The comparison results are presented in Table 2.

Table 2. Benchmark comparison for the 2024 E-Gini value.

Model	Predicted 2024	Absolute Error	Relative Error
Linear trend	0.213586	0.007014	3.180%
Univariate GM(1,1)	0.214516	0.006084	2.758%
Naive persistence	0.218103	0.002497	1.132%
ARIMA	0.219274	0.001326	0.601%
Standard SVR	0.222301	0.001701	0.771%
Proposed GM(1,1)-CS-SVR	**0.220130**	**0.000470**	**0.213%**

The results indicate that the proposed GM(1,1)-CS-SVR framework obtains the lowest forecasting error. Its forecasted E-Gini value for 2024 is 0.220130, with an absolute error of 0.000470 and a relative error of 0.213%. Among the univariate benchmarks, ARIMA performs best, with a relative error of 0.601%. The Standard SVR yields a relative error of 0.771%. These findings suggest that, under the validation design adopted in this study, the proposed hybrid framework delivers the most accurate short-term forecasting results. Moreover, compared to the univariate models, the proposed framework provides greater interpretive value by linking the forecast to substantively meaningful explanatory variables of educational inequality.

The reference data for the accuracy test grades are shown in Table 3. Moreover, the accuracy results for the three input datasets in the model are shown in Table 4. The datasets 1, 2 and 3 in Table 4 respectively represent government budget spending on education, per capita consumption expenditure on education, and CPI for education.

Table 3. Reference data for accuracy test grades.

Accuracy Grade	Mean Relative Error (%)	Grey Absolute Correlation	Mean Squared Error Ratio	Small Error Probability
First grade (excellent)	1	0.90	0.35	0.95
Second grade (good)	5	0.80	0.50	0.80
Third grade (pass)	10	0.70	0.65	0.70
Fourth grade (not applicable)	20	0.60	0.80	0.60

Table 4. Accuracy results for three input datasets.

Dataset	Mean Relative Error (%)	Grey Absolute Correlation	Mean Squared Error Ratio	Small Error Probability
1	0.912	0.9906	0.0773	1
2	0.541	0.9822	0.0204	1
3	0.298	0.9972	0.0550	1

The relative errors for all three datasets are less than 1%, the absolute grey correlations are greater than 0.9, the mean squared error ratios are less than 0.35, and the small error probabilities exceed 0.95. The fitting accuracy of all three models meets the requirements of grey theoretical modeling; therefore, it is reasonable to assume that these three datasets can be modeled as grey systems. The grey system theory model can effectively predict government budget spending on education, per capita consumption expenditure on education, and CPI for education.

Based on data from 2002 to 2023, we use the proposed framework to predict the data in 2024. The predicted 2024 values are 0.5572 ten trillion Yuan for government budget spending on education, 3248 Yuan for per capita consumption expenditure on education, and 160.8 for CPI for education. When compared with the actual data from China in

2024, government budget spending on education is 0.5416 ten trillion Yuan, per capita consumption expenditure on education is 3189 Yuan, and CPI for education is 159.57. The corresponding relative errors for these variables are 2.8807%, 1.8501%, and 0.7708%, respectively, demonstrating a satisfactory degree of accuracy.

4.2. Discussion

The proposed framework is an effective tool for short-term forecasting of educational inequality under annual small-sample conditions. Over the study period, the E-Gini of China shows an overall long-term downward trend, with stage-specific fluctuations. In this context, the results suggest that, under the validation design used in this study, incorporating government budget spending on education, per capita consumption expenditure on education, and CPI for education improves the short-term forecasting performance of the E-Gini index.

The three input variables represent different macro-level dimensions. Their combination allows the forecasting task to systematically incorporate information on public finance, household behavior, and cost pressures. The relationship between the E-Gini index and these indicators provides useful contextual information for interpreting the forecasting results.

The benchmark comparison provides evidence for evaluating the performance of the proposed framework. Among all tested models, the proposed framework yields the lowest forecasting error for the 2024 E-Gini. Its predicted value is 0.220130, with an absolute error of 0.000470 and a relative error of 0.213%. The results indicate that incorporating predicted changes in the three education-related indicators within the two-stage framework helps improve short-term forecasting performance. This performance exceeds that of univariate methods relying only on the historical E-Gini series, as well as that of the standard multivariate SVR specification. In this sense, the proposed framework provides a more informative foundation for sustainability-oriented monitoring of educational inequality.

Moreover, the findings suggest that changes in public education finance, household education expenditure, and education-related prices all deserve continued attention when assessing future changes in educational inequality. Such information helps identify emerging pressure points among public provision, household burden, and educational affordability, which are closely related to the sustainable improvement of educational equity. However, stronger claims about specific policy effects require additional causal, institutional, or distributional analysis to match the proposed framework.

5. Conclusions

In this paper, a hybrid framework for predicting trends in educational inequality is proposed. The effects of government budget spending on education, per capita consumption expenditure on education, and CPI for education are incorporated. Compared to the prediction results obtained by univariate time-series forecasting models, the results of this study have greater practical relevance. The proposed framework uses a two-step prediction strategy based on the GM(1,1) model and the CS-SVR model. First, the GM(1,1) model forecasts changes in the influencing factors. Second, based on these predictions, the CS-SVR model is used to predict the trend of educational inequality. Annual government statistics from 2002 to 2024 are used to validate the accuracy and effectiveness of the proposed framework. Moreover, the superiority of the framework is further demonstrated by comparison with the traditional SVR model. The results indicate an overall downward trend in educational inequality in China, as measured by the E-Gini index. This study offers valuable references for policymakers and researchers focused on educational inequality.

Although the methodology described above yields satisfactory results on the dataset from 2002 to 2024, the forecasting framework can be further refined due to uncertainties in future policies and other influencing factors. In addition, although the three indicators selected can capture only part of the relevant factors, many important determinants are not yet incorporated. The important factors include regional disparities in fiscal capacity and school quality, inequalities associated with the urban–rural divide and the household registration system, differences in household income and the distribution of opportunities, the expansion of shadow education, among others. It should be noted that the proposed framework has the potential to include these factors by extending the data in the first stage. In the future, the framework can be improved by incorporating additional influencing factors to enhance the reliability and robustness of the predictions.

Author Contributions: Z.G. and T.S. conceived the paper. Z.G. wrote the main manuscript text. T.S. and L.S. conducted additional review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are publicly available from official statistical sources, including the National Bureau of Statistics of China and education statistics yearbooks. The processed dataset and implementation details can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Garibaldi, P. *Personnel Economics in Imperfect Labour Markets*; Oxford University Press: Oxford, UK, 2006.
2. Galor, O.; Moav, O. From Physical to Human Capital Accumulation: Inequality and the Process of Development. *Rev. Econ. Stud.* **2004**, *71*, 1001–1026. [CrossRef]
3. Palmisano, F.; Biagi, F.; Peragine, V. Inequality of Opportunity in Tertiary Education: Evidence from Europe. *Res. High. Educ.* **2022**, *63*, 514–565. [CrossRef]
4. Song, Y.; Zhou, G. Inequality of Opportunity and Household Education Expenditures: Evidence from Panel Data in China. *China Econ. Rev.* **2019**, *55*, 85–98. [CrossRef]
5. Wu, L.; Yan, K.; Zhang, Y. Higher Education Expansion and Inequality in Educational Opportunities in China. *High. Educ.* **2020**, *80*, 549–570. [CrossRef]
6. Liu, A.; Li, W.; Xie, Y. Social Inequality in Child Educational Development in China. *Chin. J. Sociol.* **2020**, *6*, 219–238. [CrossRef]
7. Rowley, K.J.; Edmunds, C.C.; Dufur, M.J.; Jarvis, J.A.; Silveira, F. Contextualising the Achievement Gap: Assessing Educational Achievement, Inequality, and Disadvantage in High-Income Countries. *Comp. Educ.* **2020**, *56*, 459–483. [CrossRef]
8. Tesfay, N.; Malmberg, L.-E. Horizontal Inequalities in Children's Educational Outcomes in Ethiopia. *Int. J. Educ. Dev.* **2014**, *39*, 110–120. [CrossRef]
9. Dadon-Golan, Z.; BenDavid-Hadar, I.; Klein, J. Revisiting Educational (in)Equity: Measuring Educational Gini Coefficients for Israeli High Schools during the Years 2001–2011. *Int. J. Educ. Dev.* **2019**, *70*, 102091. [CrossRef]
10. Dorius, S.F. The Rise and Fall of Worldwide Education Inequality from 1870 to 2010: Measurement and Trends. *Sociol. Educ.* **2013**, *86*, 158–173. [CrossRef]
11. Thomas, V.; Wang, Y.; Fan, X. *Measuring Education Inequality: Gini Coefficients of Education*; Policy Research Working Paper No. 2525; World Bank: Washington, DC, USA, 2001.
12. Luo, G.; Zeng, S.; Baležentis, T. Multidimensional Measurement and Comparison of China's Educational Inequality. *Soc. Indic. Res.* **2022**, *163*, 857–874. [CrossRef]
13. Ding, Y. Financial Inequalities and Inequities in the Financing of Compulsory Education in China. Ph.D. Thesis, Columbia University, New York, NY, USA, 2005.

14. Davis, A.N.; Carlo, G.; Maiya, S. Towards a Multisystem, Strength-Based Model of Social Inequities in US Latinx Youth. *Hum. Dev.* **2021**, *65*, 204–216. [CrossRef]
15. Waitoller, F.R.; Maggin, D.M. Can Charter Schools Address Racial Inequities Evidenced in Access to the General Education Classroom? A Longitudinal Study in Chicago Public Schools. *Remedial Spec. Educ.* **2020**, *41*, 127–138. [CrossRef]
16. Mubarak, A.A.; Cao, H.; Zhang, W. Prediction of Students' Early Dropout Based on Their Interaction Logs in Online Learning Environment. *Interact. Learn. Environ.* **2022**, *30*, 1414–1433. [CrossRef]
17. Xu, M.; Chen, S.; Chen, J.; Zhang, T. Non-Linear Links between Human Capital, Educational Inequality and Income Inequality, Evidence from China. *PLoS ONE* **2023**, *18*, e0288966. [CrossRef] [PubMed]
18. Qian, H.; Walker, A.; Xu, X. Running Schools on Two Legs: The Impact of Policy Oscillation on a Public-Private Partnership School in China. *Int. J. Educ. Dev.* **2023**, *100*, 102806. [CrossRef]
19. Koirala, N.P.; Koirala, D.P.; Nyiwul, L.; Hu, Z. Economic Uncertainty, Households' Credit Situations, and Higher Education. *J. Macroecon.* **2024**, *80*, 103598. [CrossRef]
20. Hall, S.G.; O'Hare, B. A Model of the Impact of Government Revenue and Quality of Governance on Schooling. *Int. J. Educ. Dev.* **2024**, *108*, 103055. [CrossRef]
21. Timotheou, S.; Miliou, O.; Dimitriadis, Y.; Sobrino, S.V.; Giannoutsou, N.; Cachia, R.; Mones, A.M.; Ioannou, A. Impacts of Digital Technologies on Education and Factors Influencing Schools' Digital Capacity and Transformation: A Literature Review. *Educ. Inf. Technol.* **2023**, *28*, 6695–6726. [CrossRef]
22. Anderson, E.; Jalles D'Orey, M.A.; Duvendack, M.; Esposito, L. Does Government Spending Affect Income Inequality? A Meta-Regression Analysis. *J. Econ. Surv.* **2017**, *31*, 961–987. [CrossRef]
23. Di Gioacchino, D.; Sabani, L. Education Policy and Inequality: A Political Economy Approach. *Eur. J. Polit. Econ.* **2009**, *25*, 463–478. [CrossRef]
24. Esposito, L.; Villasenor, A. Wealth Inequality, Educational Environment and School Enrolment: Evidence from Mexico. *J. Dev. Stud.* **2018**, *54*, 2095–2118. [CrossRef]
25. Jez, S.J. The Differential Impact of Wealth Versus Income in the College-Going Process. *Res. High. Educ.* **2014**, *55*, 710–734. [CrossRef]
26. Li, L.; Bai, X.; Xia, H. Prediction of the Trend of Higher Education Development Using a Weakening Buffer Operator-Based GM (1,1) Model. *Educ. Inf. Technol.* **2024**, *29*, 2523–2538. [CrossRef]
27. Zuiderwijk, A.; Janssen, M. Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. In *Open Government: Opportunities and Challenges for Public Governance*; Gascó-Hernández, M., Ed.; Springer: New York, NY, USA, 2014; pp. 115–135.
28. Liang, J.; He, X.; Xiao, H.; Wu, C. Offshore Wind Power Prediction Based on Two-Stage Hybrid Modeling. *Energy Strateg. Rev.* **2024**, *54*, 101468. [CrossRef]
29. Zhu, Y.; Xie, C.; Sun, B.; Wang, G.-J.; Yan, X.-G. Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models. *Sustainability* **2016**, *8*, 433. [CrossRef]
30. Ziesemer, T. Global Dynamics of Gini Coefficients of Education for 146 Countries: Update to 1950–2015 and a Compact Guide to the Literature. *Bull. Appl. Econ.* **2022**, *9*, 85–95. [CrossRef]
31. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Statistics for Engineering and Information Science; Springer: New York, NY, USA, 1998.
32. Yang, Y.; Zhang, M.; Dai, Y. A Fuzzy Comprehensive CS-SVR Model-Based Health Status Evaluation of Radar. *PLoS ONE* **2019**, *14*, e0213833. [CrossRef]
33. Xu, H. Prediction of Students' Performance Based on the Hybrid IDA-SVR Model. *Complexity* **2022**, *2022*, 1845571. [CrossRef]
34. Bian, X.-Q.; Zhang, L.; Du, Z.-M.; Chen, J.; Zhang, J.-Y. Prediction of Sulfur Solubility in Supercritical Sour Gases Using Grey Wolf Optimizer-Based Support Vector Machine. *J. Mol. Liq.* **2018**, *261*, 431–438. [CrossRef]
35. Shi, T.; Pang, M.; Wang, Y.; Zhang, Y. Inverse Parameter Identification Framework for Cohesive Zone Models Based on Multi-Island Genetic Algorithm. *Eng. Fract. Mech.* **2024**, *300*, 110005. [CrossRef]
36. Cheng, P.; Wang, X. Influence of SVR Parameter on Non-Linear Function Approximation. *Comput. Eng.* **2011**, *37*, 189–194.
37. Li, X. Sequence Model and Prediction for Sustainable Enrollments in Chinese Universities. *Sustainability* **2023**, *15*, 214. [CrossRef]
38. Wang, L.-N. Prediction of Investment Supply Scale for Higher Education of China Based on Levenberg-Marquardt Algorithm. *J. Guangxi Univ.* **2009**, *34*, 635–639. [CrossRef]
39. Friedman, J.; York, H.; Graetz, N.; Woyczynski, L.; Whisnant, J.; Hay, S.I.; Gakidou, E. Measuring and Forecasting Progress towards the Education-Related SDG Targets. *Nature* **2020**, *580*, 636–639. [CrossRef]
40. Maliti, E. Inequality in Education and Wealth in Tanzania: A 25-Year Perspective. *Soc. Indic. Res.* **2019**, *145*, 963. [CrossRef]
41. Artige, L.; Cavenaile, L. Public Education Expenditures, Growth and Income Inequality. *J. Econ. Theory* **2023**, *209*, 105622. [CrossRef]

42. Chen, Y.; Yuan, M.; Zhang, M. Income Inequality and Educational Expenditures on Children: Evidence from the China Family Panel Studies. *China Econ. Rev.* **2023**, *78*, 101932. [CrossRef]
43. Muller, A. Education, Income Inequality, and Mortality: A Multiple Regression Analysis. *Br. Med. J.* **2002**, *324*, 23–25. [CrossRef] [PubMed]
44. Fan, J.; Huang, J.; Sessions, J.G.; Ye, J. Local Education Expenditures and Educational Inequality in China. *Manch. Sch.* **2023**, *91*, 283–305. [CrossRef]
45. Wei, Y. Household Expenditure on Education in China: Key Findings From the China Institute for Educational Finance Research-Household Surveys (CIEFR-HS). *ECNU Rev. Educ.* **2024**, *7*, 738–761. [CrossRef]
46. Fotso, H.R.F.; Kaze, C.V.A.; Kenmoe, G.D. A Novel Hybrid Model Based on Weather Variables Relationships Improving Applied for Wind Speed Forecasting. *Int. J. Energy Environ. Eng.* **2022**, *13*, 43–56. [CrossRef]
47. Zhang, G.P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **2003**, *50*, 159–175. [CrossRef]
48. Zhang, J.; Li, L.; Chen, W. Predicting Stock Price Using Two-Stage Machine Learning Techniques. *Comput. Econ.* **2021**, *57*, 1237–1261. [CrossRef]
49. Liu, J.; Xiao, X.; Guo, J.; Mao, S. Error and Its Upper Bound Estimation between the Solutions of GM(1,1) Grey Forecasting Models. *Appl. Math. Comput.* **2014**, *246*, 648–660. [CrossRef]
50. Yang, Y.; Feng, J. Fault Pattern Recognition and State Prediction Research of Ship Power Equipment Based on HMM-SVR. *Ship Eng.* **2018**, *40*, 68–72+97.
51. Sulastri, H.; Intani, S.M.; Rianto, R. Application of Bagging and Particle Swarm Optimisation Techniques to Predict Technology Sector Stock Prices in the Era of the COVID-19 Pandemic Using the Support Vector Regression Method. *Int. J. Comput. Sci. Eng.* **2023**, *26*, 255–267. [CrossRef]
52. Zhang, J.; Teng, Y.-F.; Chen, W. Support Vector Regression with Modified Firefly Algorithm for Stock Price Forecasting. *Appl. Intell.* **2019**, *49*, 1658–1674. [CrossRef]
53. Rajabioun, R. Cuckoo Optimization Algorithm. *Appl. Soft Comput.* **2011**, *11*, 5508–5518. [CrossRef]
54. Yang, X.-S.; Deb, S. Cuckoo Search via Lévy Flights. In Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009; pp. 210–214.
55. Yang, X.-S.; Deb, S. Multiobjective Cuckoo Search for Design Optimization. *Comput. Oper. Res.* **2013**, *40*, 1616–1624. [CrossRef]
56. Jin, C.; Jin, S.-W. Software Reliability Prediction Model Based on Support Vector Regression with Improved Estimation of Distribution Algorithms. *Appl. Soft Comput.* **2014**, *15*, 113–120. [CrossRef]
57. Nie, J. The Gini Coefficients of Education: Evidence from China. *Pop. Dev.* **2006**, *12*, 42–47.
58. Wu, F.; Zhang, Y.; Wu, Y. The Influence of Higher Continuing Education on Educational Equality in China. *Mod. Distance Educ. Res.* **2023**, *35*, 68–74.
59. Huang, W.; Chen, N.; Zhang, X. The Fluctuation of Educational Inequality and the Education Expansion and Demographic Transition—The Kuznets Curve of Education since the Founding of New China. *Res. Educ. Tsinghua Univ.* **2019**, *40*, 55–63.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.