| Problem Chosen | 2021 MCM/ICM Summary Sheet | Team Control Number |
|---|---|---|
| F | | 2126835 |

# Health and Sustainability Evaluation Model for Nation's Higher Education System Based on GA-BP Neural Network and Time Series Analysis


## Summary

In the past 40 years, the development of higher education has been widely concerned around the world, which has played a negligible roll in all society fields. In this paper, a suite of models to easily and directly assess the health and sustainability level of any nation's higher education system are built, in view of both current condition and development tendency. The higher education systems of Brazil, China, Japan, South Africa and the United States are evaluated as examples.

Firstly, 10 indicators involved in three aspects of Equity, Investment and Quality are selected to help build the health and sustainability evaluation model, particularly, we define the geographical distribution with Lorenz Curve and Gini Index as analog to discuss the relationship among higher education equity, administrative region and population. By using Principal Component Analysis (PCA) method, the ten indicators are processed and simplified into 3 principal components, with each representing one of the three aspects. Furthermore, Back Propagation (BP) Neural Network is introduced into the model with the 3 principle components as input layer nodes and a total evaluation level as the output layer node. For lack of reality data, 1500 random number groups signed with a grading mechanism is used as the standard training set. Though with the advantage of rich data tolerance, performance of BP neural network shows slow processing speed and low precision. The further improvement of the neural network model by using of Genetic Algorithm helps gain better performance in convergence. So far we gain a successfully trained evaluation model together with the evaluated health level of the five nations' higher education system.

Secondly, in this paper, time series analysis is used to predict the future of higher education of the five nations in the next three years. The ARIMA model was introduced to predict changes in the next three years based on three indicators for five countries within five years. Then the prediction results are input to the health evaluation model, so that the health status of the higher education system of each country three years later can be inferred, which can be regarded as its sustainability.

Thirdly, this paper takes Brazil's higher education system as an case study. Based on the analysis using the above model, Brazil is rated as Level 1 in both health and sustainability, with considerable room for improvement. Considering with the student-to-staff ratio and employment rate, which have the greatest impact on the results, and combining with the current economic background and the status quo of higher education system in Brazil, we put forward several effective reform measures, which can raise the health level of Brazil's higher education system to 2 or even 3, with obvious effects. This prediction result is based on the previously constructed model. Of course, in reality, the nation's industrial structure, existing education system and concepts will all have an impact on the prediction of the effect of measures, which will be discussed in this paper.

Finally, we analyze the performance and the sensitivity of the model, which prove that our model is stable under different situations.

# Contents

# 1 Introduction

## 1.1 Problem Background

A system of higher education is an important element in a nation's efforts to further educate its citizens beyond required primary and secondary education, and therefore has value both as an industry itself and as a source of trained and educated citizens for the nation's economy. In the past 40 years, the expansion of higher education has become a mainstream trend in the development of higher education around the world, which has had a profound impact on all dimensions of the society and the higher education system. Worldwide, an increasing proportion of people are enrolled in higher education [1]. According to data compiled by UNESCO, the global gross enrollment rate (GTER) in higher education increased from 10% in 1972 to 32% in 2012, and is now growing by 1% a year. By 2014, 64 countries had a gross enrollment ratio of 50 percent in higher education, compared with only five countries 20 years ago, and 14 countries globally had a gross enrollment ratio of more than 80 percent in higher education [2]. Scholars represented by Martin Trow have studied the education in the stage of popularization of higher education, believing that higher education in the stage of popularization will shape a new relationship among the state, education and society, and that the internal characteristics of higher education will also change [3]. In the process of popularization of higher education, these predictions have been confirmed and revised constantly. The changes in the higher education system in the process of universalization is of great theoretical significance and practical value to clarify this problem for promoting the sustainable and healthy development of higher education.

## 1.2 Literature Review

Up to now, the United Nations and other international organizations have not systematacially conducted statistics and evaluation for higher education. But historically, the United Nations' assessment of universal education has served as a guide. The Education Index, one of the three components of the Human Development Index (HDI) published by the United Nations Development Programme, is measured by adult literacy rates (2/3 weight) and overall enrolment rates of primary, secondary and university (1/3 weight) [4].

## 1.3 Our work

1. We first develop several models that can assess the health and sustainability of any nation's system of higher education.

2. Then we apply the models to five representative countries. We select a nation whose system of higher education has room for improvement and propose an attainable and reasonable vision that supports a healthy and sustainable system of higher education.

3. We use the models to measure the health of both the current system and proposed, healthy, sustainable system for the selected nation. And we propose targeted policies and an implementation timeline that will support the migration from the current state to your proposed state, and use the models to shape and/or assess the effectiveness of your policies. Then we discuss the real-world impacts of implementing the plan both during the transition and in the end state.

4. Finally, we analyze the performance and the sensitivity of the model to prove that our model is stable under different situations.

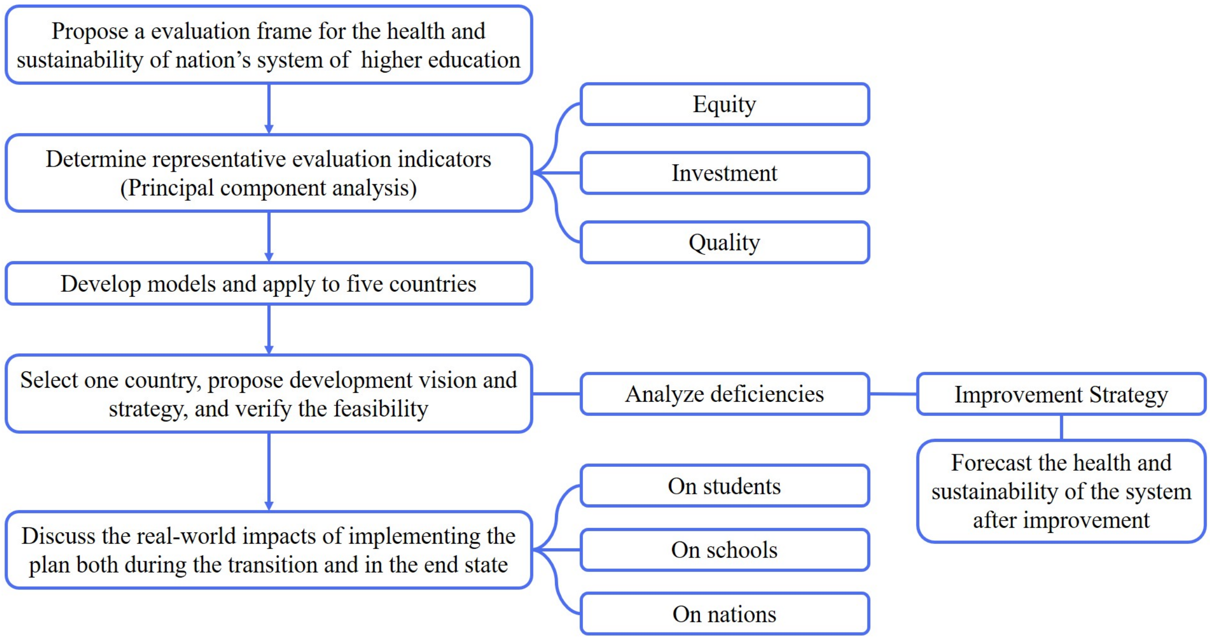



Figure 1: Flow chart of our work

# 2 Assumptions

In our models mentioned later, we present the following five assumptions:

**1.** assume that the education data collected are reliable in all countries;

**2.** assume that the 10 factors mentioned in the paper are used to represent all the factors that may affect higher education;

**3.** assume that the five countries mentioned in the paper are used to represent the countries to be analyzed;

**4.** assume that the analysis process is not affected by extreme data, and it will not occur that the absence of one index in a university will lead to an excessively distorted evaluation result in the overall evaluation of the whole country;

**5.** assume that in the process of research and testing of key principal components, only the influence of the change of key principal components on the evaluation results is taken into account, and the interaction between affecting principal components is temporarily excluded.

# 3 Model Establishment and Solution

## 3.1 Frame Description

In this section, we construct evaluation models for both health and sustainability of higher education in different countries, and use the model to evaluate. Based on the results of the evaluation, we make suitable suggestions for countries with room for improvement in higher education.

## 3.2 Indicators of Evaluation Models for National Higher Education

### 3.2.1 Indicator Selection

In this paper, we establish 3 aspects and 10 indicators (shown as Table 1) to measure the health and sustainability of each nation's higher education system.

Table 1: Indicators of Evaluation Models for National Higher Education

| Aspects | Indicators |
|---|---|
| Equity | Geographical distribution |
| | Enrollment rate |
| | Gender ratio |
| Investment | Education expenditure-to-GDP ratio |
| | Staff-to-student ratio |
| | Proportion of capital expenditure |
| Quality | Average number of SCI papers per staff |
| | Employment rate |
| | Proportion of international students |
| | Occupation ratio of QS Top 200 |

Equity aspect shows the overall situation of higher education in the country. This not only reflects the government's attention to the popularization of higher education, but also reflects the public's attitude and understanding towards higher education. Investment aspect is the measurement of the efforts and achievements made by the government and universities in education and research. Quality aspect explains the substantive results achieved by the higher education system, and we use several indicators to reflect its overall level of research output, student quality, and international influence. Next, we explain what each indicator means and the data is shown in Appendix.

- Geographical distribution (Table 9 and 10): is reflected by the degree to which the population of each province matches the number of higher education institutes, which will be explained by the Gini index, and it will be described in more detail below.
- Enrollment rate (Table 11): is calculated as a ratio of the number of registered students to the school-age population.
- Gender ratio (Table 12): is determined by the ratio of the number of boys to girls in higher education institutes, which is a direct reflection of gender equality in the higher education system, which is a reflection of the country's citizens' access to higher education.
- Education expenditure-to-GDP ratio (Table 13): is the proportion of government spending on education in the current year's Gross Domestic Product (GDP), which shows the importance and investment of the country to the higher education system.
- Student-to-staff ratio (Table 14): is obtained by the ratio of the number of students to that of teachers in higher education system, which reflects the allocation of teachers' resources and the investment of the universities.
- Proportion of capital expenditure (Table 15): is determined by the proportion of capital expenditure to the total expenditure on higher education. This proportion reflects the importance attached to campus construction, which is necessary to ensure normal teaching and research.

- Average number of SCI papers per staff (Table 16): is calculated by the ratio of the number of SCI papers produced by the nation's higher education system to the number of teachers in that year, which objectively reflects the scientific research capability of the system.
- Employment rate (Table 17): is calculated by the proportion of fresh graduates who are employed within six months, which is representative of the students' professional quality.
- Proportion of international students (Table 18): is calculated by the proportion of international students in a country's higher education system as a percentage of total students, which reflects international recognition.
- Occupation ratio of QS Top 200 (Table 19): is the percentage of the nation's universities in Top 200 of the QS World University Rankings that year, which reflects international reputation.

**Geographical Distribution by Administrative Region** It is carried out in similar way with income equality analysis of Gini Index, the ranking of income per person is replaced by ranking of ratio between university number in an administrative region and population of the region (in which way ignore the quality difference between universities but focus on numbers only).

The Gini Index is strictly linked to the representation of income inequality through the Lorenz Curve [5]. In particular, it measures the ratio of the area between the Lorenz Curve and the equidistribution line (henceforth, the concentration area) to the area of maximum concentration. In this geographical distribution model,

$$q_i = y_1 + y_2 + y_3 + \cdots + y_i \tag{1}$$

$$p_i = n_1 + n_2 + n_3 + \cdots + n_i \tag{2}$$

where $y_i$ is university number per ten thousand people by administrative region in ascending order, $n_i$ is population ratio of total in the same order, $q_i$ is cumulative proportion of university number and $p_i$ is cumulative proportion of population. Therefore, the Gini index of geographical distribution is expressed as

$$G = 1 - \sum_i [(q_i + qi - 1)(p_i - pi - 1)] \tag{3}$$

and considering that absolute average is not desired, we set 0.2 as the ideal number for the Gini index.

### 3.2.2 Dimensionless Processing and Normalization

Based on the 3 aspects and 10 indicators, we analyzed the data of one country in one year as a sample. In multivariate statistical analysis, the collected data are of different dimensions, and there are differences in orders of magnitude or units of measurement between them, so the variables are not consistent. If the data are directly used, the model will be greatly affected by the magnitude of some data, or the calculation results will not be of practical significance due to the disunity of measurement units. Therefore, we need to standardize the data first, which is dimensionless processing. In this way, the dimensionality and order of magnitude of each indicator can be eliminated to ensure the smooth progress of the model. Then we

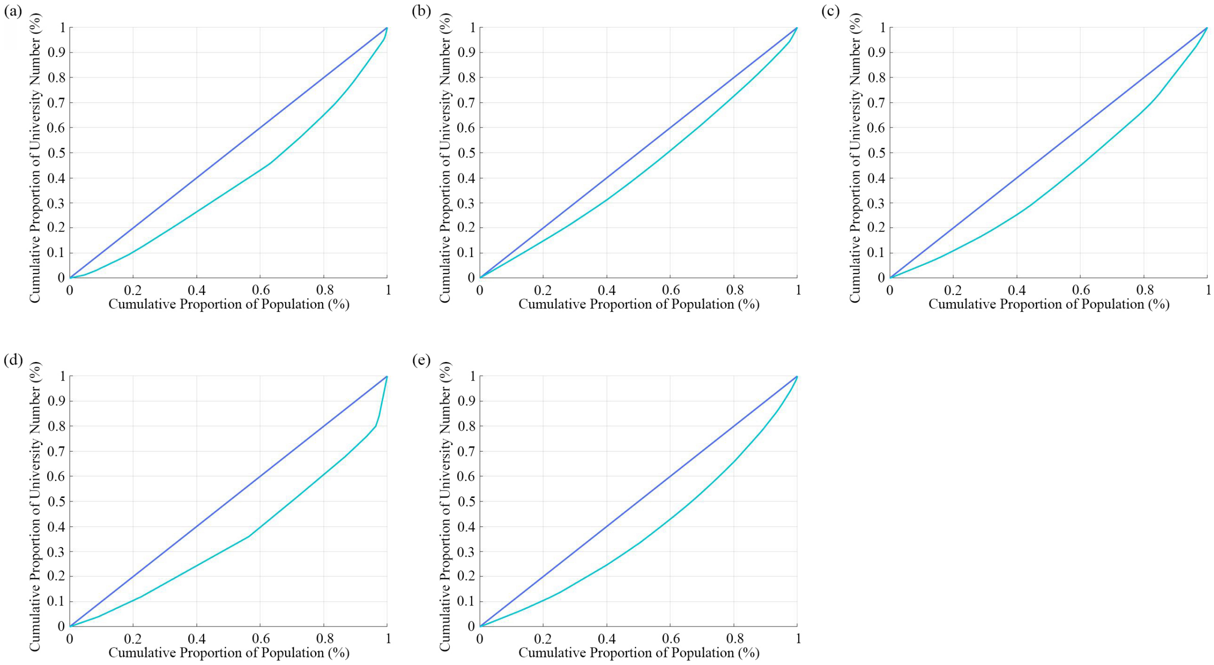



Figure 2: Gini Index for the proportion of university number and population in (a) Brazil, (b) China, (c) Japan, (d) South Africa, and (e) United States

normalized the data and converted the original data to the range of [0,1] using linear functions. The normalization formula is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{min} - X_{max}} \tag{4}$$

where $X_{norm}$ is the data after normalization, $X$ is the original data, and $X_{max}$ and $X_{min}$ are the maximum and minimum values of the original data set. In this way, equal scaling of the original data is achieved.

#### 3.2.3 Principal Component Analysis to Reduce the Dimension of Indicators

For the 10 indicators of a sample quantified above, if the data is directly analyzed, it will be affected by noise and extreme data. At the same time, the convergence speed of the model will be seriously affected by too much data in the input layer. Therefore, it is necessary for us to adjust and reduce dimensions of all indicators. In this paper, the indicators have been classified into different regions, so it is only necessary to extract the principal components of the indicators in each region.Since the extraction has nothing to do with the year, a total of 25 samples of data from 5 countries in 5 years were directly analyzed by principal component analysis.

**Suitability Test** We set $r(X_i, X_j)$ as the correlation coefficient of $X_i$ and $X_j$, then

$$r\left(X_i, X_j\right) = \frac{Cov\left(X_i, X_j\right)}{\sqrt{Var\left[X_i\right] Var\left[X_j\right]}} \tag{5}$$

where $Cov\left(X_i, X_j\right)$ is the covariance of $X_i$ and $X_j$, $Var\left[X_i\right]$ is the variance of $X_i$, and $Var\left[X_j\right]$ is the variance of $X_j$.

We can judge whether principal component analysis can be used according to the correlation coefficient of the indicators in each aspect. By calculating, we found that the average correlation coefficient $r(X_i, X_j)$ of Equity, Investment, and Quality aspects are 0.63, 0.76, and 0.85, which are not less than 0.3, so that the principal component analysis is feasible.

**Principal Analysis** According to the cumulative contribution rate, in order to avoid excessive loss of the data information contained in the princess component and improve the utility of principal component analysis, if the cumulative variance contribution rate of principal components is greater than 80%, these can be selected as the extraction result of principal components. Finally, we extracted the first principal component respectively, and they represent variables in three aspects, Equity, Investment, and Quality. These three principal components (as shown in Table 2) can be considered as factors for the evaluation of national higher education.

Table 2: Three Principal Components Extracted from Equity, Input and Quality Aspects

| Component | Year | Country | | | | |
|---|---|---|---|---|---|---|
| | | Brazil | China | Japan | South Africa | United States |
| Equity | 2014 | -0.0347 | 0.2454 | -0.5663 | 0.7927 | -0.4370 |
| | 2015 | -0.0241 | 0.1997 | -0.5786 | 0.8163 | -0.4134 |
| | 2016 | 0.0103 | 0.1666 | -0.6005 | 0.8260 | -0.4023 |
| | 2017 | 0.0063 | 0.1529 | -0.6301 | 0.8480 | -0.3771 |
| | 2018 | -0.0238 | 0.2028 | -0.6285 | 0.8137 | -0.3643 |
| Investment | 2014 | -0.4982 | -0.07190 | 0.9935 | -0.3291 | -0.09437 |
| | 2015 | -0.6975 | -0.02493 | 0.9419 | -0.0880 | -0.1315 |
| | 2016 | -0.5383 | -0.3809 | 0.8785 | 0.02286 | 0.01776 |
| | 2017 | -0.5780 | -0.1782 | 0.8747 | -0.07454 | -0.04406 |
| | 2018 | -0.6051 | -0.1285 | 0.8725 | -0.1424 | 0.003490 |
| Quality | 2014 | 0.6634 | 0.8931 | -0.2839 | -0.4365 | -0.8360 |
| | 2015 | -0.6626 | -0.8454 | 0.2500 | 0.3675 | 0.8906 |
| | 2016 | -0.6495 | -0.7760 | 0.2185 | 0.2914 | 0.9156 |
| | 2017 | -0.7058 | -0.6788 | 0.2227 | 0.2565 | 0.9055 |
| | 2018 | -0.7136 | -0.6060 | 0.2803 | 0.1082 | 0.9310 |

## 3.3 Health Evaluation Models for Nation's Higher Education System

### 3.3.1 Overview of Model Selection, Training, and Application

In this part, we set up an evaluation system based on BP neural network according to the three main factors extracted in the previous part. Due to the lack of a training set, we will try to use standard data for training. After obtaining the model, we will evaluate the health of higher education in five countries. In the evaluation, we graded the five countries on a level-by-level basis with ideal health status in higher education as Level 5 and least ideal health status as a Level 1.

### 3.3.2 Back Propagation (BP) Neural Network

Back Propagation (BP) neural network (Fig. 3) is a kind of multi-layer feedforward neural network [7], its main characteristics are signal forward transmission and error back propagation.

In forward transmission, input signals are processed layer by layer from the input layer through the hidden layer to the output layer. The state of neurons in each layer only affects the state of neurons in the next layer. If the output layer is not able to obtain the expected output, it will turn to back propagation and adjust the network weights and thresholds according to the prediction errors, so that the predicted output of BP neural network is constantly approaching the expected output.

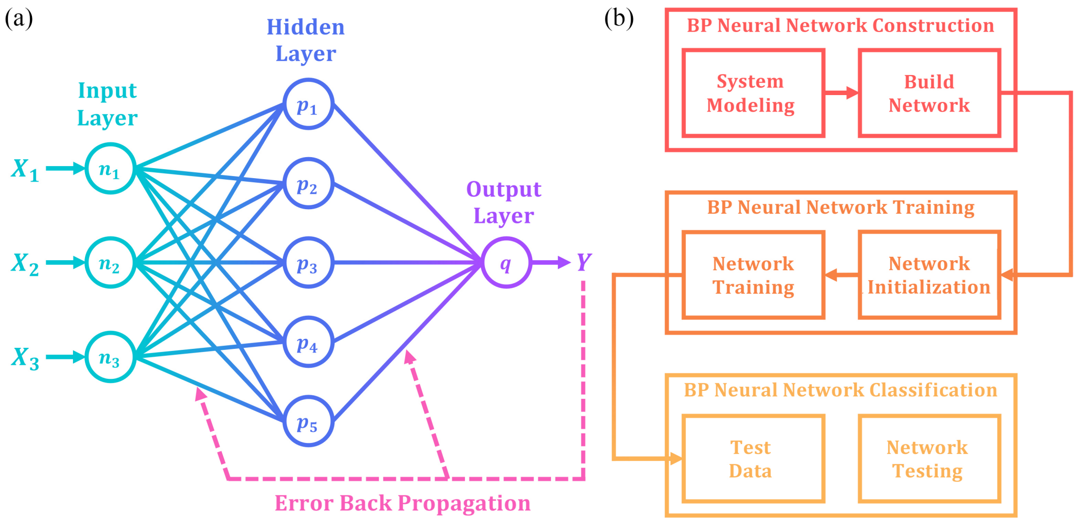



Figure 3: Overview of the Back Propagation (BP) Neural Network: (a) the structure of the BP neural network; (b) the flow chart of the BP neural network construction, training and classification

**Feasibility Test** First of all, we need to test the correlation of the factors to verify the feasibility of using BP neural network. We let $r\left(X_i, X_j\right)$ be the correlation coefficient between indicators and using Eq. 5 to calculate. According to Table 3, most of $r\left(X_i, X_j\right)$ are not less than 0.2, it indicates that the three factors can be used as the input layer of the neural network, that is, BP neural network is feasible.

Table 3: Correlation Test Results for Three Components

| Component | Equity | Investment | Quality |
|---|---|---|---|
| Equity | 1.000 | -0.566 | -0.355 |
| Investment | -0.566 | 1.000 | 0.508 |
| Quality | -0.355 | 0.508 | 1.000 |

According to the requirements of the problem, we set the input layer to contain 3 nodes, the hidden layer to contain 5 nodes, and the output layer to contain 1 node. We set up a series of samples as a training set based on the behaviorally anchored rating scale, which covers several samples of different levels to reflect the diversity.

**Constraints for Standard Training Set** The three components Equity(E), Investment(I), Quality(Q) are scored with -2, -1, 0, 1, 2 in different value ranges. And for the changing value compared with former year $\Delta E$, $\Delta I$, $\Delta Q$, only their signs are taken into consideration.

The total evaluated score of a country in a certain year mainly comes from scores of E, I and Q, together with some evaluation on the organic connection [6] between $E$, $I$, $Q$, $\Delta E$, $\Delta I$ and $\Delta Q$ as shown in Table 4. Despite reference from 'Handbook on Measuring Equity in Education', an incentive mechanism is also introduced: when a certain component (E, I or Q) is low but with rising tendency would gain extra scores.

The standard training set are composed of 1500 random number groups and signed by hand according to the constraints above (sign their final scores or evaluation levels).

Table 4: Influence Coefficient between Principal Components

| | $\Delta E<0$ | $\Delta I<0$ | $\Delta Q<0$ | $\Delta E>0$ | $\Delta I>0$ | $\Delta Q>0$ |
|---|---|---|---|---|---|---|
| High $E$ | 0 | 0 | 0 | 0 | 0 | 0 |
| High $I$ | - | 0 | 0 | 0 | 0 | 0 |
| High $Q$ | 0 | 0 | 0 | 0 | 0 | 0 |
| Low $E$ | 0 | 0 | 0 | + | 0 | 0 |
| Low $I$ | 0 | 0 | 0 | 0 | + | 0 |
| Low $Q$ | 0 | - | 0 | 0 | + | + |

**BP Neural Network Training** First of all, according to the problems, we define the network input layer node number $n$ is 3, the hidden layer node number $l$ is 5, the output layer node number $m$ is 1, initialize the connection weights between the input layer and hidden layer $\omega_{ij}$ and that between the hidden layer and output layer $\omega_{jk}$, initialize the hidden layer threshold $a$, the output layer threshold $b$, and give the learning rate $\eta$ and the excitation function $f$.

According to the input vector $X$, the connection weight $\omega_{ij}$ between the input layer and the hidden layer, and the hidden layer threshold $a$, the hidden layer output $H$ is calculated as

$$H_j = f(\sum_{i=1}^{n} \omega_{li} x_i - a_j) \quad j = 1,2,\cdots,l \tag{6}$$

where $l$ is the number of nodes in the hidden layer, and $f$ is the excitation function. The function $f$ can be expressed in a number of ways, and we choose

$$f(x) = \frac{1}{1+c^x} \tag{7}$$

According to the hidden layer output $H$, the connection weight $\omega_{jk}$ and the threshold $b$, the predicted output $O$ can be calculated as

$$O_k = \sum_{j=1}^{l} H_j \omega_{jk} - b_k \quad k = 1,2,\cdots,m \tag{8}$$

Then according to the network forecast output $O$ and the expected output $Y$, the network forecast error $e$ can be calculated as

$$e_k = Y_k - O_k \quad k = 1,2,\cdots,m \tag{9}$$

According to the network prediction error $e$, the network connection weights $\omega_{ij}$ and $\omega_{jk}$ can be updated by using Eq. 10 and 11.

$$\omega_{ij} = \omega_{ij} + \eta\, H_j(1-H_j)x(i) \sum_{k=1}^{m} \omega_{jk} e_k \quad i = 1,2,\cdots,n; j = 1,2,\cdots,l \tag{10}$$

$$\omega_{jk} = \omega_{jk} + \eta\ H_j e_k \quad j = 1,2,\cdots l; k = 1,2,\cdots,m \tag{11}$$

where $\eta$ is learning rate.

According to the network prediction error $e$, the threshold $a$ and $b$ can be updated by using Eq. 12 and 13.

$$a_j = a_j + \eta\ H_l(1 - H_j)\sum_{k=1}^{m} \omega_{jk} e_k \quad j = 1,2,\cdots,l \tag{12}$$

$$b_k = b_k + e_k \quad k = 1,2,\cdots,m \tag{13}$$

After the completion of network update, judge whether the threshold value is reached: if it is, then the process ends and the trained model is obtained; if not, the iterative process is repeated until the requirement is met.

### 3.3.3 Improvement of Neural Network by Introducing Genetic Algorithm

In the process of testing BP neural network, we found that the network was not stable, and sometimes there would be too long convergence time or a big difference between weight and threshold value in each test. This shows that in this problem, pure BP neural network is difficult to get a good evaluation effect. In general, the determination of network weights and thresholds has always been a difficulty in network training, and the randomly selected parameters will seriously affect the accuracy of the network. Therefore, we consider to add genetic algorithm into the neural network, and use the optimal individuals trained by the genetic algorithm to optimize the weight and threshold of the network.

First, we conduct individual coding and population initialization. The individual contains the ownership values and thresholds of the entire neural network. In this paper, individuals are coded with real numbers. The encoding length formula is:

$$S = n \times m + m \times l + m + l \tag{14}$$

where, $S$ is the encoding length, $m$ is the number of nodes of hidden layers, $n$ is the number of nodes in the input layer, $l$ is the number of nodes in the output layer. The size of the population has a great impact on the global search performance of the genetic algorithm, and the size of the initial population we selected is 20.

Next, we set the fitness function as the reciprocal of the sum of squares of neural network errors:

$$f = \frac{1}{SE} \tag{15}$$

where $SE$ is the sum of the squares of errors between the predicted and expected outputs of the neural network.

Then we select individuals, which can be selected according to the probability value, and the formula is as follows:

$$P_i = \frac{f_i}{\sum_{i=1}^{k} f_i} \tag{16}$$

where is $f_i$ individual fitness value, and $k$ is the number of individuals in the population.

Then we do the crossover and the mutation. The optimal individual does not cross over, but is copied directly into the next generation. Other individuals will be mutated using the probability of mutation $p_m$ to produce another new individual. In this experiment, the probability of mutation $p_m$ is 0.7, the probability of crossover $p_c$ is 0.3, and the evolutionary algebra is selected as 100. This process is repeated until a result is obtained (Fig. 4).

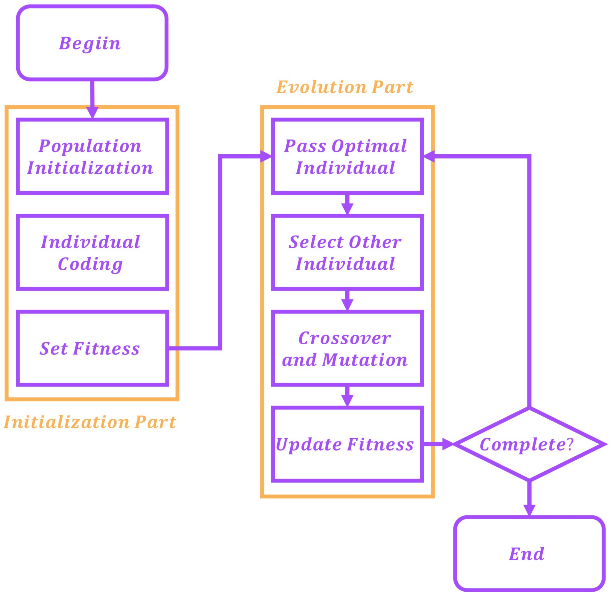



Figure 4: Flow chart of the genetic algorithm

The weights and thresholds optimized by the genetic algorithm will be used in the BP neural network. The following steps are consistent with those described in Section 3.3.2. After training, the error of GA-BP neural network (Table 5) is within a reasonable range, and the model is effective. Therefore, we can use this model to evaluate the health of nation's higher education systems.

### 3.3.4 Level Standard for the Health Evaluation Models

As mentioned in Section 3.3.2, we explained the score standard and constraint for the three components and how to get the final scores of one nation in one year. It gave us the final scores, which will be used to determine the level. We use the behaviorally anchored rating scale (BARS) to determine the corresponding score range for each level (Table 5). This standard has been repeatedly tested and compared, and can be considered to be capable of distinguishing the levels of different nation's higher education systems.

Table 5: Level Standard for the Health Evaluation Models

| Level | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Score | $< -1$ | $-1 \sim 0$ | $0 \sim 1$ | $1 \sim 2$ | $> 2$ |

### 3.3.5 Results of the Health Evaluation Models for Nation's Higher Education System

We used the GA-BP neural network to get the evaluation of the higher education systems in five countries, as shown in the Table 6.

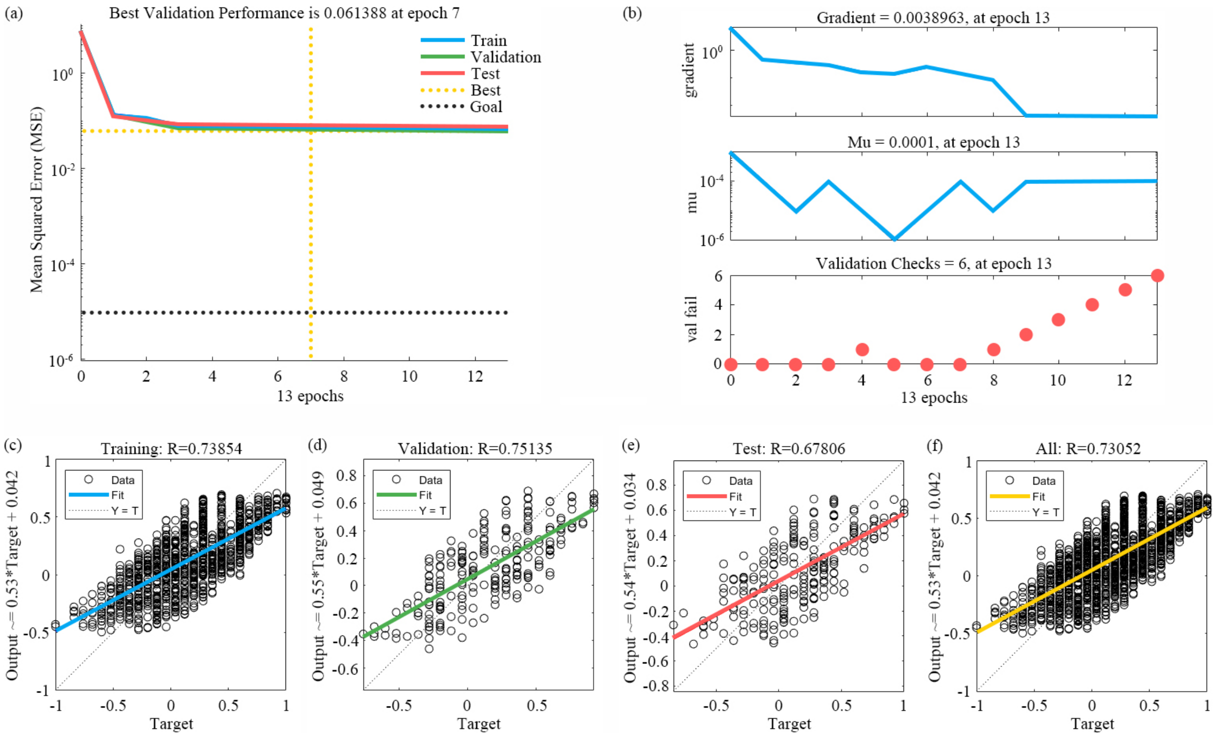



Figure 5: Error Analysis for GA-BP neural network: (a) best validation performance, (b) gradient, mu and validation checks, and output error and fitted curve for (c) training data, (d) validation data, (e) test data, and (f) all data

Table 6: Results of the Genetic Algorithm-Back Propagation (GA-BP) Neural Network

| | Brazil | China | Japan | South Africa | United States |
|---|---|---|---|---|---|
| 2015 | -2.2 | -1 | 3.2 | 1 | 2.3 |
| 2016 | -1.8 | -1.7 | 3 | 1.1 | 2.8 |
| 2017 | -2 | -0.9 | 3 | 0.8 | 2.6 |
| 2018 | -2.1 | -0.7 | 3.1 | 0.2 | 2.8 |

According to the standard in Table 5, we can clearly see the current health of higher education systems in each nation. Brazil is worst, with a health level of 1; China has a health level of 2; Japan is the best, and the sustainability level is 5; South Africa is slipping, but still with a health level of 3; the U.S. is strong, and the health level is 5.

## 3.4 Sustainability Evaluation Models for Nation's Higher Education System

### 3.4.1 Overview of Model Selection, Training, and Application

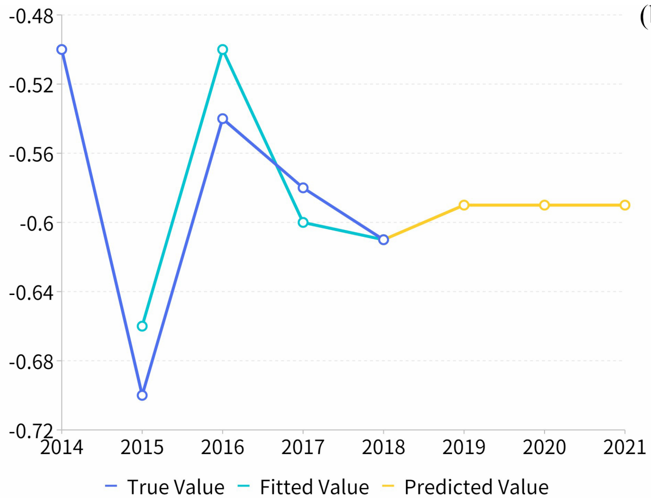
The sustainable evaluation of national higher education is the evaluation of its development in the future. We can make time series analysis based on the existing multi-year data to predict the future development, and then apply the health evaluation model, GA-BP neural network, to evaluate the predicted results to reflect the sustainability.

### 3.4.2 Time Series Analysis

We use time series analysis based on the ARIMA model. The model requires stationarity of sequence. Therefore, we conducted ADF test. When the difference $d$ is of order 1, the significance $p$ value is 0, significantly less than 0.01, indicating that the sequence meets the stationarity. Thus, the national higher education system satisfies the ARIMA (1,1,1) model.

Now we take Brazil's Investment Component as an example to make time series analysis (Fig. 6). We applied the ARIMA model to obtain a number of parameters (Fig. 6 (b)), and obtained the fitting values and predicted values in the next three years, as shown in Fig. 6 (a), according to the prediction equation. Q statistic can be used to test white noise. In this case, the magnitude of Q6 is 0.312, greater than 0.1, which indicates that the residual of the model is white noise and the model meets the requirements.

(a)




(b)

| ARIMA(1, 1, 1) Parameter List | | |
|---|---|---|
| Parameter | Symbol | Value |
| Constant | $c$ | $-0.59$ |
| AR | $\alpha 1$ | $-0.641$ |
| MA | $\beta 1$ | $-1$ |
| Q Statistic | $Q6$ ($p$ Value) | $0.312(0.577)$ |
| | $Q12$ ($p$ Value) | $null(null)$ |
| | $Q18$ ($p$ Value) | $null(null)$ |
| | $Q24$ ($p$ Value) | $null(null)$ |
| | $Q30$ ($p$ Value) | $null(null)$ |
| Information Criteria | $AIC$ | $-11.227$ |
| | $BIC$ | $-12.79$ |

Figure 6: Results of time series analysis for investment component of Brazil in next three years: (a) the line chart which contains true, fitted and predicted values, and (b) model parameter list

After repeating the above steps, we obtained the stationary time series of the three components of the samples from all countries and all years, and predicted the changes of each Component in the next three years (Table 7).

Table 7: Predictions of Factors for Five Countries in the Next Three Years Based on Time Series Analysis

| | Brazil | | | China | | | Japan | | | South Africa | | | United States | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2019 | 2020 | 2021 | 2019 | 2020 | 2021 | 2019 | 2020 | 2021 | 2019 | 2020 | 2021 | 2019 | 2020 | 2021 |
| Equity | -0.03 | -0.02 | -0.02 | 0.21 | 0.19 | 0.18 | -0.61 | -0.60 | -0.60 | 0.80 | 0.83 | 0.82 | -0.39 | -0.40 | -0.39 |
| Investment | -0.59 | -0.59 | -0.59 | -0.16 | -0.16 | -0.16 | 0.89 | 0.91 | 0.92 | 0.12 | -0.12 | -0.12 | -0.05 | -0.02 | -0.02 |
| Quality | -0.67 | -0.65 | -0.64 | -0.54 | -0.46 | -0.40 | 0.26 | 0.22 | 0.20 | 0.10 | 0.03 | 0.01 | 0.89 | 0.95 | 0.92 |

### 3.4.3 Results of the Sustainability Evaluation Models for Nation's Higher Education System

Finally, we substituted the predicted results into the GA-BP network mentioned in Section 3.3, and the results (Table 8) reflected the sustainability of national higher education.

According to the standard in Table 5, we can clearly see the development trend of the higher education system in each nation in the next three years. Brazil remains in the doldrums, with a sustainability level of 1; China has improved slightly, with a sustainability level of 2; Japan remains stable and good situation, and the sustainability level is 5; South Africa will slip slowly,

Table 8: Results of the GA-BP Neural Network Based on Forecast Three-Year Data

| | Brazil | China | Japan | South Africa | United States |
|---|---|---|---|---|---|
| 2019 | -1.9 | -0.7 | 3 | 0.2 | 3 |
| 2020 | -1.9 | -0.6 | 3 | 0 | 2.7 |
| 2021 | -1.9 | -0.6 | 3.1 | 0 | 2.7 |

with a sustainability level of 3; the U.S. will rise slightly in a choppy, and the sustainability level is 5.

## 3.5 Nation's Higher Education System Improvement: A Case Study of Brazil

Based on the evaluation results of the health and sustainability of the higher education system of the five countries in Section 3.3 and 3.4, we can find that the higher education system of Brazil is both Level 1, with a large room for improvement. Now, we will analyze the current situation, improvement strategies and practical challenges of health and sustainability according to the models analysis results and indicators of statistical data.

### 3.5.1 Analysis of the Existing Situation

Combining the results in Table 6 and 8 and relevant data in Appendix, the most outstanding problem in higher education system is that although with a great proportion of GDP devoted into tertiary education and a relatively fair geographical distribution, many critical indicator are still terrible. However, with a command of tertiary education actuality in Brazil, these data may make sense: their public tertiary education are small in amount but great in quality , while the private tertiary education takes in most of students but too poor in quality. The public part consumes most of the education investment in GDP, but not contributes to general tertiary education level. The private part improves the geographical distributing condition, but poor quality provides no help to the critical indicators.

### 3.5.2 Targeted Policies and Implementation Timeline

Policies should focus on narrowing the gap between public and private higher education by slightly leaning the investment towards private ones and control the progress of popularization to ensure education quality.

**Student-to-staff ratio** Brazil's student-to-staff ratio is close to 20:1, which leads to a low score. A low number of teachers leads to a lack of resources and inadequate academic guidance for students, which affects the quality of education. The government can expand the number of teachers by encouraging talents to work as teachers in universities and actively introducing overseas teachers, as well as adopt government regulations to control the adverse influence of the disproportionate number of students enrolled by colleges and universities compared with the number of teachers recruited. In this respect, we suggest that Brazil can improve its student-to-staff ratio to 12:1, which can be obtained by the health evaluation model calculation as -0.8, and the evaluation level is upgraded to 2, which is obviously effective.

**Employment Rate** The employment rate directly reflects the teaching quality of higher education. With only 50% of graduates employed in Brazil, the large number of jobless graduates is a drag on social stability and economic prosperity. The government can take policies such as expanding infrastructure construction to provide more jobs and give companies incentives to hire college graduates.. Universities should pay more attention to the employability of graduates, increase the development of application-oriented skills in daily teaching activities, and provide students with more internship opportunities. In this respect, if Brazil can make the employment rate increase to 80%, the health evaluation score will be -1.0, and the evaluation level will be 2.

All of the improvements mentioned above were incorporated into the model using the single-variable principle, where only one value was changed. If multiple improvement strategies are adopted at the same time, the improvement effect will be more obvious. When the student-to-staff ratio and employment rate are both increased, the model calculates that Brazil's health and sustainability score will be close to 0, which means that the evaluation level will be close to Level 3, which is significant. Moreover, the increase in employment will depend not only on the actions of the government and universities, but also on the number of jobs available in the industry itself. Brazil's current economic structure is a serious constraint on industry and, without reform, will continue to weigh on employment.

#### 3.5.3 Challenges For Reform of Higher Education System in Reality

Due to Brazil's own economic downturn, higher education spending, though large, is still difficult to fully support the large number of private colleges and the expensive spending for public institutions. What's more, private schools are likely to hire fewer teachers because of the economic benefits of the current trend to expand enrollment to meet their profits, which is in conflict with a higher student-to-staff ratio.

# 4 Model Strengths and Weaknesses

## 4.1 Model Strengths

- The model and data in this paper make a general survey of the global higher education situation from both horizontal and vertical dimensions, and select nations at different levels for analysis, which are representative.
- 10 representative indicators are selected to reflect the situation of nation's higher education system, combined with the data from several rich and authoritative database, such as World Bank Database, UNESCO Institute of Statistics Database and national statistical yearbooks.
- The Gini index is introduced into the calculation of the geographical distribution, which makes the data more convincing. And the indicator, geographical distribution, is different from the common evaluation model, which provides a new method to investigate the important scale of fairness.
- The combination of genetic algorithm and neural network has high classification accuracy, strong robustness and fault-tolerant ability, and can fully approximate the complex nonlinear relationship.

## 4.2 Model Weaknesses

- In the past year, the world has suffered from the impact of the COVID-19, and the economy of all countries is in the gloom. Online higher education is promoted for the first time and the immature response methods have caused the noise in last year's data, which will not disappear in the next two years. At the same time, the noise data is scarce, which is difficult to be used to predict the future development, so that the model has limitations.
- Due to the lack of country-specific and professional-specific data on higher education, it is not possible to conduct analysis and empirical research on specific majors.
- In time series analysis, the number of sample durations is relatively short, and the predicted results may have large deviations.s

# 5 Model Improvement and Promotion

## 5.1 Model Improvement

### 5.1.1 Optimize the Number of Hidden Layer Nodes of BP Neural Network

The number of nodes in the hidden layer of BP neural network has a great influence on the prediction accuracy of BP neural network. With too many nodes, the training time increases, and it is easy to overfit. The following formula can be used for the selection of the optimal number of hidden layer nodes:

$$l < n - 1 \tag{17}$$

$$l < \sqrt{(m-n)} + a \tag{18}$$

$$l = log_2 n \tag{19}$$

where $n$ is the number of nodes in the input layer, $l$ is the number of nodes of hidden layer, $m$ is the number of nodes in the output layer, and $a$ is a constant between 0 and 10. In practical problems, the selection of hidden layer node number is firstly determined by referring to the formula to determine the approximate range of node number. Then the best number of nodes is determined by trial and error method. For some problems, the number of hidden layer nodes has little influence on the output results.

### 5.1.2 Attach the Momentum Method

The BP neural network adopts gradient correction method as the learning algorithm of weight and threshold, and corrects the weight and threshold from the negative gradient direction of network prediction error, without considering the accumulation of previous experience, and the learning process converges slowly. For this problem, the momentum method can be attached to solve it. The weight learning formula with additional momentum is:

$$\omega(k) = \omega(k-1) + \Delta\omega(k) + a[\omega(k-1) - \omega(k-2)] \tag{20}$$

where $\omega(k)$, $\omega(k-1)$ and $\omega(k-2)$ is the weight when the time is $k$, $k-1$ and $k-2$, and $a$ is the momentum learning rate.

### 5.1.3 Adaptive learning rate algorithm

The learning rate $\eta$ of BP neural network is between [0,1]. The larger the learning rate $\eta$ is, the greater the weight modification is, and the faster the network learning speed is. However, if the learning rate $\eta$ is too large, the weight will oscillate in the learning process, and if the learning rate is too small, the network convergence will be too slow and the weight will not be stable. The adaptive learning rate method means that the learning rate $\eta$ is large in the early stage of BP neural network evolution, and the network converges rapidly. With the progress of learning process, the learning rate decreases continuously and the network tends to be stable. One of the most efficient adaptive learning rate algorithm, Adam algorithm (Algorithm 1), designs independent adaptive learning rate for different parameters by calculating the first and second moment estimation of gradient. After bias correction, the learning rate of each iteration has a fixed range, which makes the parameters relatively stable. It takes into account both speed of convergence and stationarity.

**Algorithm 1** The Adam Algorithm

**Require:** Step size $\varepsilon$ (Suggested default: 0.001)
**Require:** Exponential decay rates for moment estimates, $\rho_1$ and $\rho_2$ in $[0,1)$. (Suggested defaults: 0.9 and 0.999 respectively)
**Require:** Small constant $\sigma$ used for numerical stabilization. (Suggested defaults: $10^{-8}$)
**Require:** Initial parameters $\theta$

Initialize 1st and 2nd moment variables $s=0$, $r=0$
Initialize time step $t=0$
**while** stopping criterion not met **do**
  Sample a minibatch of $m$ examples from the training set $\{x^{(1)},\ldots,x^{(m)}\}$ with
  corresponding target $y^{(i)}$.
  Compute gradient: $g \leftarrow \frac{1}{m}\nabla_\theta \sum_i L(f(\theta^{(i)};\theta), y^{(i)}$
  $t \leftarrow t+1$
  Update biased first moment estimate: $s \leftarrow \rho_1 s + (1-\rho_1)g$
  Update biased second moment estimate: $r \leftarrow \rho_2 r + (1-\rho_2)g \odot g$
  Correct bias in first moment: $\hat{s} \leftarrow \frac{s}{1-\rho_1^t}$
  Correct bias in second moment: $\hat{r} \leftarrow \frac{r}{1-\rho_2^t}$
  Compute update: $\Delta\theta = -\varepsilon\frac{\hat{s}}{\sqrt{\hat{r}}+\delta}$ (operations applied element-wise)
  Apply update: $\theta \leftarrow \theta + \Delta\theta$
**end while**

## 5.2 Model Promotion

The health and sustainable evaluation models established in this work can be evaluated not only in higher education, but also in various fields, such as politics, economy, military, agriculture, etc. [8]. The evaluation results are more accurate and objective and have more reference value by combining the short-term and long-term evaluation system.

# 6 Model Sensitivity Analysis

## 6.1 Sensitivity Analysis for Health Evaluation Model

In order to test the sensitivity of the health evaluation model, we chose the Monte Carlo algorithm to perturb the initial data of each country in each year, set the deviation of 1%, 3% and 5% respectively, and then use the previously trained health evaluation model to evaluate. After calculation, the final results are 94.3%, 91.5% and 88.4% the same as the original results respectively, which indicates that the model still has a strong stability after the data is interfered.

## 6.2 Sensitivity Analysis for Sustainability Evaluation Model

In order to test the sensitivity of the sustainability evaluation model, we also chose the Monte Carlo algorithm to perturb the initial data of each nation in each year, set the deviation of 1%, 3% and 5% respectively, and then use the time series analysis to get the forecast data and finally input them to the previously trained health evaluation model to evaluate the sustainability. After calculation, the final results are 95.1%, 93.9% and 86.0% the same as the original results respectively, which indicates that the whole model still has a strong stability after the data is interfered.

# References


[1] Jinyuan Ma,Yuzhuo Cai. Innovations in an institutionalised higher education system: the role of embedded agency. Higher Education,2021(prepublish).

[2] Nicolai Götze,Teresa Carvalho,Timo Aarrevaara. Academics' Societal Engagement in Diverse European Binary Higher Education Systems: A Cross-Country Comparative Analysis. Higher Education Policy,2021(prepublish).

[3] Mahmoudi Fahimeh,Bagheri Majd Rouhollah. The effect of lean culture on the reduction of academic corruption by the mediating role of positive organizational politics in higher education. International Journal of Educational Development,2021,80.

[4] YunTao Pan. Research on The Quality of Graduate Education from The Perspective of Whole Process Quality Management. International Journal of Social Sciences in Universities,2020,3(4).

[5] Bellù, Lorenzo Giovanni and Liberati, Paolo. Inequality Analysis: The Gini Index. EASYPol Module 040,2003.

[6] UNESCO Institute of Statistics. Handbook on Measuring Equity in Education. 2018, from `http://uis.unesco.org/sites/default/files/documents/handbook-measuring-equity-education-2018-en.pdf`

[7] Song Gao,Juan Huang,Yaru Li,Guiyan Liu,Fan Bi,Zhipeng Bai. A forecasting model for wave heights based on a long short-term memory neural network. Acta Oceanologica Sinica,2021(prepublish).

[8] Michael Kariwo,Tatiana Gounko,Musembi Nungu. A Comparative Analysis of Higher Education Systems - Issues, Challenges and Dilemmas[M].Brill Sense:2014-01-01.

# Appendix: Data Set of Evaluation Models

In this appendix, we list all the raw data that we used in Section 3. These data are from World Bank databases ( `https://data.worldbank.org/`), UNESCO databases ( `http://data.uis.unesco.org/`, `https://www.whed.net/`), and reports from the National Bureau of Geography and Statistics of Brazil ( `https://www.ibge.gov.br/`), the National Bureau of Statistics of China ( `http://www.stats.gov.cn/`), and the National Center of Education Statistics of USA ( `https://nces.ed.gov/`).

Table 9: Geographical Distribution of Higher Education Institutes and Population by Provinces in Each Nation

**United States**

| Administrative Region | HEI Number | Population |
|---|---|---|
| Alabama | 37 | 4,934,190 |
| Alaska | 5 | 724,357 |
| Arizona | 23 | 7,520,100 |
| Arkansas | 21 | 3,033,950 |
| California | 184 | 39,613,500 |
| Colorado | 31 | 5,893,630 |
| Connecticut | 25 | 3,552,820 |
| Delaware | 5 | 990,334 |
| District of Columbia | 12 | 714,153 |
| Florida | 79 | 21,944,600 |
| Georgia | 59 | 10,830,000 |
| Hawaii | 9 | 1,406,430 |
| Idaho | 7 | 1,860,120 |
| Illinois | 92 | 12,569,300 |
| Indiana | 44 | 6,805,660 |
| Iowa | 37 | 3,167,970 |
| Kansas | 27 | 2,917,220 |
| Kentucky | 31 | 4,480,710 |
| Louisiana | 25 | 4,627,000 |
| Maine | 19 | 1,354,520 |
| Maryland | 29 | 6,169,040 |
| Massachusetts | 87 | 6,912,240 |
| Michigan | 63 | 9,992,430 |
| Minnesota | 45 | 5,706,400 |
| Mississippi | 16 | 2,966,410 |
| Missouri | 54 | 6,169,040 |
| Montana | 12 | 1,085,000 |
| Nebraska | 20 | 1,952,000 |
| Nevda | 9 | 3,185,790 |
| New Hampshire | 16 | 1,372,200 |
| New Jersey | 30 | 8,874,520 |
| New Mexico | 13 | 2,917,220 |
| New York | 171 | 19,300,000 |
| North Carolina | 56 | 10,701,000 |
| North Dakota | 13 | 770,026 |
| Ohio | 72 | 11,714,600 |
| Oklahoma | 26 | 3,990,440 |
| Oregon | 29 | 4,289,440 |
| Pennsylvania | 129 | 12,804,100 |
| Puerto Rico | 48 | 3,194,370 |
| Rhode Island | 10 | 1,061,510 |
| South Carolina | 33 | 5,277,830 |
| South Dakota | 13 | 896,581 |
| Tennessee | 54 | 6,944,260 |
| Texas | 94 | 29,730,300 |
| Utah | 13 | 3,310,770 |
| Vermont | 20 | 623,251 |
| Virginia | 52 | 8,603,980 |
| Washington | 31 | 7,796,940 |
| West Virginia | 23 | 1,767,860 |
| Wisconsin | 45 | 5,852,490 |
| Wyoming | 2 | 581,075 |

**Brazil**

| Administrative Region | HEI Number | Population |
|---|---|---|
| DISTRITO FEDERAL | 5 | 2,852,372 |
| ACRE | 2 | 790,101 |
| ALAGOAS | 5 | 3,321,730 |
| AMAPA | 3 | 750,912 |
| AMAZONAS | 5 | 750,912 |
| ESPIRITO SANTO | 7 | 3,885,049 |
| BAHIA | 11 | 15,126,371 |
| CEARÁ | 8 | 8,842,791 |
| GOIÁS | 5 | 6,523,222 |
| MARANHÃO | 4 | 6,850,884 |
| MATO GROSSO | 7 | 6,850,884 |
| MATO GROSSO DO SUL | 5 | 2,619,657 |
| MINAS GERAIS | 20 | 20,734,097 |
| PARÁ | 8 | 8,073,924 |
| PIAUÍ | 4 | 3,194,178 |
| PARAÍBA | 6 | 3,943,885 |
| PARANÁ | 16 | 11,081,692 |
| PERNAMBUCO | 3 | 9,277,727 |
| RIO DE JANEIRO | 22 | 16,461,173 |
| RIO GRANDE DO SUL | 24 | 11,207,274 |
| RIO GRANDE DO NORTE | 5 | 3,408,510 |
| RONDÔNIA | 3 | 1,748,531 |
| RORAIMA | 3 | 496,936 |
| TOCANTINS | 3 | 1,496,880 |
| SANTA CATARINA | 14 | 6,727,148 |
| SÃO PAULO | 44 | 44,035,304 |

**China**

| Administrative Region | HEI Number | Population |
|---|---|---|
| Beijing | 93 | 19612368 |
| Tianjin | 56 | 12938693 |
| Hebei | 122 | 71854210 |
| Shanxi | 82 | 35712101 |
| Inner Mongolia | 53 | 24706291 |
| Liaoning | 115 | 43746323 |
| Jilin | 62 | 27452815 |
| Heilongjiang | 81 | 38313991 |
| Shanghai | 64 | 23019196 |
| Jiangsu | 167 | 78660941 |
| Zhejiang | 108 | 54426891 |
| Anhui | 120 | 59500468 |
| Fujian | 90 | 36894217 |
| Jiangxi | 103 | 44567797 |
| Shandong | 146 | 95792719 |
| Henan | 141 | 94029939 |
| Hubei | 128 | 57237727 |
| Hunan | 125 | 65700762 |
| Guangdong | 154 | 104320459 |
| Guangxi | 78 | 46023761 |
| Hainan | 20 | 8671485 |
| Chongqing | 65 | 28846170 |
| Sichuan | 126 | 80417528 |
| Guizhou | 72 | 34748556 |
| Yunnan | 81 | 45966766 |
| Tibet | 7 | 3002165 |
| Shaanxi | 95 | 37327379 |
| Gansu | 49 | 25575263 |
| Qinghai | 12 | 5626723 |
| Ningxia | 19 | 6301350 |
| Xinjiang | 54 | 21815815 |

**Japan**

| Administrative Region | HEI Number | Population |
|---|---|---|
| Tokyo | 126 | 13,513,734 |
| Kanagawa | 27 | 9,127,323 |
| Osaka | 50 | 8,838,908 |
| Aichi | 47 | 7,484,094 |
| Saitama | 27 | 7,261,271 |
| Chiba | 28 | 6,224,027 |
| Hyōgo | 35 | 5,536,989 |
| Hokkaido | 34 | 5,383,579 |
| Fukuoka | 30 | 5,102,871 |
| Shizuoka | 11 | 3,701,181 |
| Ibaraki | 9 | 2,917,857 |
| Hiroshima | 18 | 2,844,963 |
| Kyoto | 30 | 2,610,140 |
| Miyagi | 14 | 2,334,215 |
| Niigata | 16 | 2,305,098 |
| Nagano | 7 | 2,099,759 |
| Gifu | 12 | 2,032,533 |
| Tochigi | 9 | 1,974,671 |
| Gunma | 12 | 1,973,476 |
| Okayama | 16 | 1,922,181 |
| Fukushima | 8 | 1,913,606 |
| Mie | 7 | 1,815,827 |
| Kumamoto | 9 | 1,786,969 |
| Kagoshima | 6 | 1,648,752 |
| Okinawa | 7 | 1,434,138 |
| Shiga | 6 | 1,413,184 |
| Yamaguchi | 11 | 1,405,007 |
| Ehime | 5 | 1,385,840 |
| Nagasaki | 8 | 1,377,780 |
| Nara | 10 | 1,365,008 |
| Aomori | 9 | 1,308,649 |
| Iwate | 5 | 1,279,814 |
| Ōita | 5 | 1,166,729 |
| Ishikawa | 11 | 1,154,343 |
| Yamagata | 4 | 1,122,957 |
| Miyazaki | 14 | 1,104,377 |
| Toyama | 4 | 1,066,883 |
| Akita | 5 | 1,022,839 |
| Kagawa | 4 | 976,756 |
| Wakayama | 3 | 963,850 |
| Yamanashi | 11 | 835,165 |
| Saga | 2 | 833,245 |
| Fukui | 4 | 787,099 |
| Tokushima | 4 | 756,063 |
| Kōchi | 3 | 728,461 |
| Shimane | 2 | 694,188 |
| Tottori | 2 | 573,648 |

**South Afica**

| Administrative Region | HEI Number | Population |
|---|---|---|
| Eastern Cape | 4 | 6508000 |
| Free State | 2 | 2891000 |
| Gauteng | 6 | 14661000 |
| KwaZulu-Natal | 4 | 1121500 |
| Limpopo | 2 | 5854000 |
| Mpumalanga | 1 | 452300 |
| North West | 1 | 3925000 |
| Northern Cape | 1 | 1230000 |
| Western Cape | 4 | 6650000 |

Table 10: Gini Index for Geographical Distribution of Higher Education Institutes and Population by Provinces in Each Nation

| | Brazil | China | Japan | South Africa | United States |
|---|---|---|---|---|---|
| Gini Index* | 0.235671 | 0.130366 | 0.216729 | 0.291344 | 0.238032 |

* Gini index is calculated by the method in Section 3.2.1 based on the data in Table 9.

Table 11: Enrollment rate of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 49.98% | 51.57% | 50.28% | 51.34% | 51.82% |
| China | 42.43% | 46.04% | 48.02% | 49.07% | 50.60% |
| Japan | 77.32% | 78.32% | 79.31% | 81.34% | 82.65% |
| South Africa | 21.80% | 22.90% | 23.00% | 23.50% | 24.20% |
| United States | 88.63% | 88.89% | 88.84% | 88.17% | 88.30% |

Table 12: Gender ratio of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 0.7452 | 0.7526 | 0.7518 | 0.7593 | 0.7574 |
| China | 0.9566 | 0.9430 | 0.9326 | 0.9326 | 0.9275 |
| Japan | 1.1307 | 1.1187 | 1.1069 | 1.0965 | 1.0895 |
| South Africa | 0.7210 | 0.7195 | 0.7240 | 0.7094 | 0.6933 |
| United States | 0.7761 | 0.7782 | 0.7835 | 0.7728 | 0.7672 |

Table 13: Education expenditure-to-GDP ratio of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 1.15% | 1.34% | 1.38% | 1.52% | 1.54% |
| China | 4.10% | 4.26% | 5.22% | 1.08% | 1.34% |
| Japan | 0.75% | 0.71% | 0.63% | 0.64% | 0.65% |
| South Africa | 0.74% | 0.75% | 0.81% | 0.88% | 0.94% |
| United States | 1.36% | 1.37% | 1.21% | 1.45% | 1.41% |

Table 14: Student-to-staff ratio of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 18.94 | 19.31 | 19.46 | 19.23 | 19.50 |
| China | 17.68 | 17.73 | 17.07 | 17.81 | 17.56 |
| Japan | 14.12 | 14.30 | 14.46 | 14.53 | 14.63 |
| South Africa | 25.71 | 26.52 | 26.74 | 26.25 | 26.95 |
| United States | 12.46 | 12.35 | 12.24 | 12.02 | 11.93 |

Table 15: Proportion of Capital Expenditure of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 7.9647 | 4.3606 | 4.9093 | 2.6497 | 2.8459 |
| China | 15.2749 | 16.2677 | 17.2451 | 18.2632 | 18.7322 |
| Japan | 18.9882 | 14.8693 | 10.3995 | 10.0823 | 10.1437 |
| South Africa | 7.2545 | 6.2752 | 5.6823 | 5.1495 | 4.7264 |
| United States | 10.8643 | 9.9418 | 4.6621 | 9.7073 | 8.3671 |

Table 16: Average number of SCI papers per staff of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 0.123 | 0.125 | 0.130 | 0.130 | 0.134 |
| China | 0.111 | 0.118 | 0.126 | 0.146 | 0.157 |
| Japan | 0.192 | 0.191 | 0.190 | 0.189 | 0.190 |
| South Africa | 0.176 | 0.179 | 0.180 | 0.185 | 0.185 |
| United States | 0.204 | 0.208 | 0.214 | 0.218 | 0.220 |

Table 17: Employment rate of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 57.1% | 56.1% | 55.8% | 53.8% | 54.3% |
| China | 92.1% | 91.7% | 90.6% | 91.9% | 91.5% |
| Japan | 67.1% | 65.4% | 66.3% | 65.1% | 63.6% |
| South Africa | 40.2% | 41.6% | 42.3% | 43.1% | 43.9% |
| United States | 53.3% | 54.0% | 57.9% | 56.5% | 56.1% |

Table 18: Proportion of international students of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 0.24% | 0.24% | 0.24% | 0.24% | 0.24% |
| China | 0.26% | 0.28% | 0.31% | 0.36% | 0.40% |
| Japan | 3.44% | 3.43% | 3.73% | 4.27% | 4.73% |
| South Africa | 4.18% | 4.12% | 4.28% | 4.06% | 3.59% |
| United States | 4.28% | 4.65% | 5.04% | 5.18% | 5.21% |

Table 19: Occupation ratio of QS Top 200 of Each Nation's Higher Education System

| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| Brazil | 0.50% | 0.50% | 1.00% | 1.00% | 1.00% |
| China | 7.00% | 7.00% | 7.50% | 7.50% | 7.00% |
| Japan | 4.50% | 4.50% | 4.00% | 4.00% | 4.50% |
| South Africa | 0.50% | 0.50% | 0.50% | 0.50% | 0.50% |
| United States | 25.50% | 25.50% | 24.50% | 24.00% | 23.50% |