Research on Prediction and Evaluation Methods for Olympics Host Countries and Cities

Yilin Wang [1], Haotian Chu [2], Heng Wu [2]

[1] College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China

[2] College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China

Abstract. Aiming at problem: By reviewing the literature and the application criteria for the right to host the Olympic Games, we selected GDP and total national reserves as economic indicators, value added of services as service level indicators, national population and national prime age labor force ratio as social indicators, national land ratio as land use indicators, and birth mortality rate as medical level indicators to establish evaluation criteria for hosting the Olympic Games, and through linear regression to The influence factor analysis of each indicator was conducted, and the significant degree of influence of each indicator was obtained, among which the influence of labor force, GDP and total reserves was more significant. By judging the influence factors, we selected GDP, total national reserves, value added of service industry, national population, proportion of prime labor force, and birth mortality rate as the indicators to continue evaluating the hosting of the Olympic Games, and selected the minimum value of each indicator as the conditional threshold through quantitative analysis to establish the threshold coefficient matrix. According to the conditions of Olympic Games submission bid, the data of each indicator in the future 2022-2029 are predicted by VAR model, and then the country or city hosting the Olympic Games is predicted whether it can be held normally by BP neural network and the conditional threshold coefficient matrix. If there is no country or city that meets the conditions for holding the Olympic Games, we choose the United States as the fixed host city and use the data from 2012-2020 to predict the index data for the next 8 years by Markov for example. Therefore, textual solution has certain feasibility.

Keywords: Multiple linear regression; Markov prediction; VAR model; BP neural network; Olympic Games.

1. Problem Background

The International Olympic Committee (IOC) is currently confronted with a diminishing number of bids to host the Olympics – both the Summer and Winter Games. Historically, the privilege of hosting the Olympics was associated with intense competition and prestige. Nevertheless, recent host cities/nations have encountered an array of adverse consequences, both short-term and long-term. In order to tackle these challenges, forward-thinking problem-solvers are contemplating various alternatives and strategies. One possibility is to establish a permanent venue for both the Summer and Winter Games. Additionally, there is a proposal to divide Olympic sports into four distinct categories, instead of the current two, thereby organizing four smaller Olympic Games (Winter, Spring, Summer, and Fall). This proposed system could potentially alleviate the burden associated with hosting such a grand-scale event.

2. The Establishment and Solution Of Problem Model

We utilized data from the International Statistical Yearbook to assess the significance of the aforementioned gross domestic product (GDP), total national reserves, service sector value-added, national population, prime labor force proportion, national land distribution, and healthcare quality indicators as independent variables. To ensure that these variables respond sensitively to dependent variable changes, the data were normalized to the same order of magnitude, enabling a more accurate evaluation of the impact factors influencing the hosting of the Olympic Games. The coefficients of the independent variables were then regressed to establish a reasonably sized model.

Multiple linear regression employs several primary influencing factors as independent variables to explain variations in the dependent variable. In real-world problem analysis, dependent variable changes are often influenced by multiple significant factors, necessitating the use of two or more factors as independent variables to account for the dependent variable's variations. This approach is known as multiple regression. When multiple independent variables exhibit a linear relationship with the dependent variable, the conducted regression analysis is referred to as multiple regression.

Let Y be the dependent variable representing whether to host the Olympics or not, and $x_{i1}, x_{i2} \cdots x_{ik} i = 1, 2 \cdots k$ be the independent variables representing the seven influencing factors of GDP, national total reserves, value added of services, national population, national prime labor force share, national land share, and health care level, respectively. When the relationship between the independent and dependent variables is linear, the multiple linear regression model is:

$$Y = b_{i0} + b_{i1}x_{i1} + \cdots + b_{ik}x_{ik}$$

where b_{i0} is the constant term, $b_{i1}, b_{i2}, \ldots, b_{ik}$ is the regression coefficient, b_{i1} is the effect of each unit increase in x_{i1} on Y when $x_{i1}, x_{i2}, \ldots, x_{ik}$ is fixed, i.e., the bias regression coefficient of x_{i1} on Y; similarly, b_{i2} is the effect of each unit increase in $x_{i1}, x_{i2}, \ldots, x_{ik}$ is fixed, the effect of each unit increase of x_{i2} on Y, i.e., the partial regression coefficient of x_{i2} on Y, etc.

Hendricks (1996) introduced Mean Relative Bias (MRB) and Root Mean Squared Relative Bias (RMSRB) as measures to objectively evaluate various VAR models based on historical data by calculating the difference in outcomes. If T represents the sample size and n is the number of VAR models, the expressions for the i-th VAR model's MRB and RMSRB are as follows:

$$MRB_i = \frac{1}{T}\sum_{t-1}^{T}\frac{VAR_{it} - \overrightarrow{VAR_t}}{\overrightarrow{VAR_t}}$$

$$RMSRB_i = \sqrt{\frac{1}{T}\sum_{t-1}^{T}\left(\frac{VAR_{it} - \overrightarrow{VAR_t}}{\overrightarrow{VAR_t}}\right)^2}$$

Among them, VAR_{it} represents the value of the i-th VAR model at time t, $\overrightarrow{VAR_t} = \frac{1}{N}\sum_{i=1}^{N} VAR_{it}$ represents the mean value of the estimated values of N VAR models at time t. Both MRB and RMSRB are dimensionless and both are independent of the unit of measure used by the VAR model. MRB reflects the degree of deviation of a VAR model estimate from the average value of all VAR model estimates, and RMSRB reflects the variability of each VAR model estimate in the form of mean square error. A small value indicates that its estimate is closer to the average. Forecasting of indicator data for the next 8 years through the VAR model:

Table 1. two of autocorrelation coefficients

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	1.000000	-0.654004	0.973775	0.307290	0.371089	0.683140
y_2	-0.654004	1.000000	-0.618022	-0.159111	-0.123346	-0.620283
y_3	0.973775	-0.618022	1.000000	0.201683	0.251995	0.603084
y_4	0.307290	-0.159111	0.201683	1.000000	0.347568	0.716702
y_5	0.371089	-0.123346	0.251995	0.347568	1.000000	0.480868
y_6	0.683140	-0.620283	0.603084	0.716702	0.480868	1.000000

Two different sets of data were predicted by adjusting different model parameter.

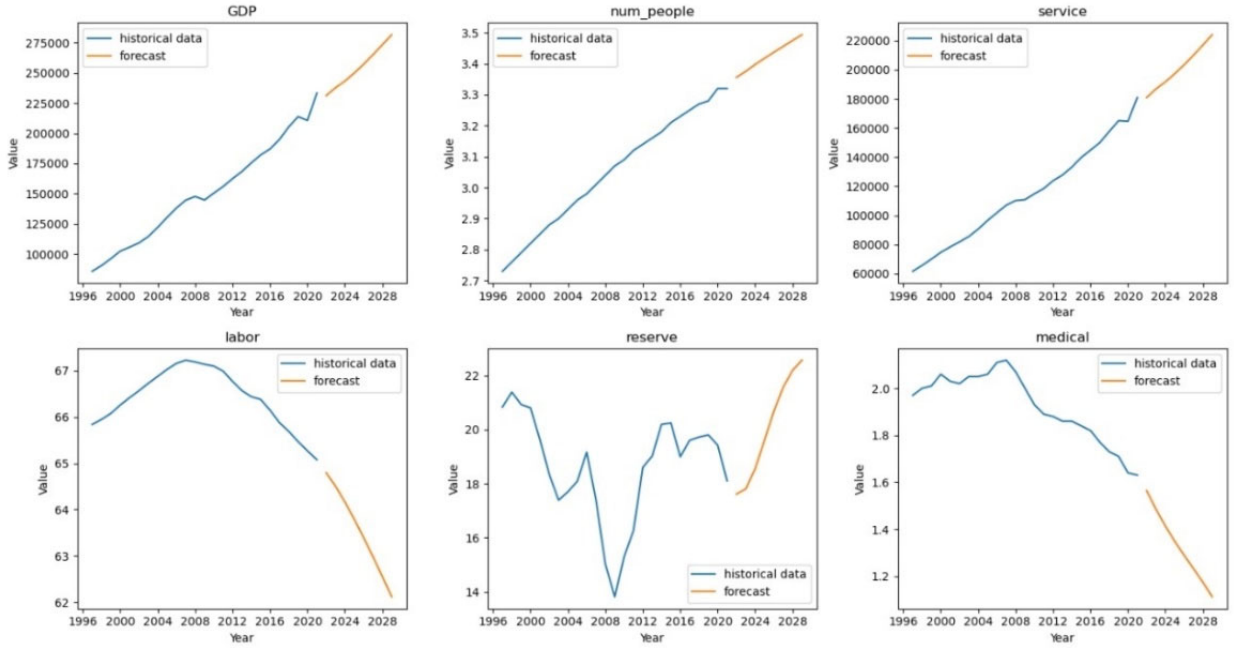

Fig. 1 The effect of prediction results of each index

Table 2. four of Forecast values of each indicator for the next 8 years

	GDP	num_people	service	labor	reserve	medical
2022	220330.860273	3.358706	173222.331605	64.698877	20.511697	1.563915
2023	258033.700587	3.355187	198482.451781	64.558718	18.115520	1.530225
2024	229826.501501	3.401754	182529.078151	64.071982	21.947829	1.487231
2025	279720.454404	3.391158	215745.988180	63.970713	17.417724	1.421090
2026	237136.170671	3.445432	190569.543442	63.346515	22.595122	1.387754
2027	305551.440112	3.421776	235122.823603	63.300584	16.958010	1.301821
2028	243947.457398	3.490006	198179.393556	62.555612	24.383338	1.277303
2029	335435.178587	3.450104	257504.436168	62.602960	16.352279	1.185327

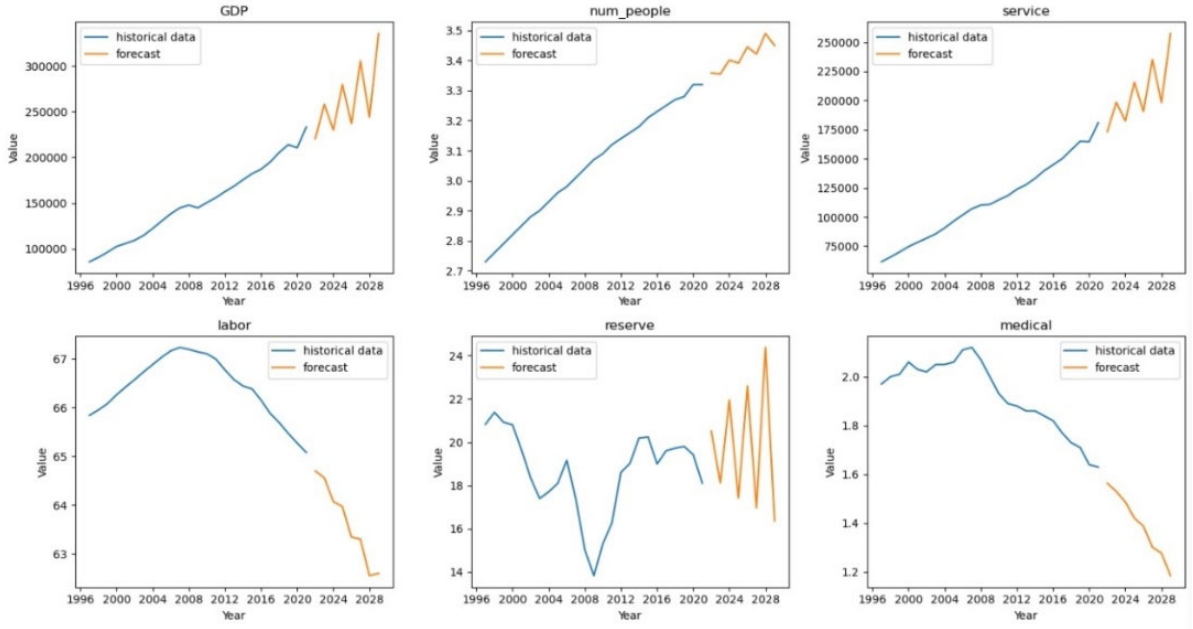

Fig. 2 The effect of prediction results of each index

Table 3. five of Forecast values of each indicator for the next 8 years

	GDP	num_people	service	labor	reserve	medical
2022	231126.053984	3.355863	180891.325845	64.797454	17.612699	1.563992
2023	237703.823328	3.375545	186843.899846	64.506621	17.804626	1.484761
2024	243191.092837	3.397439	191677.862007	64.166772	18.548523	1.413486
2025	249841.792856	3.417527	197292.878713	63.794624	19.600716	1.348198
2026	257117.326247	3.437053	203364.868522	63.397265	20.672351	1.288921
2027	265000.489444	3.456051	209956.867297	62.983285	21.562227	1.232259
2028	273226.160060	3.474874	216909.483995	62.557149	22.184999	1.174484
2029	281577.254508	3.493723	224067.534710	62.119720	22.564083	1.112416

In this paper, we used a Back Propagation (BP) neural network to predict the feasibility of hosting the Olympic Games based on two datasets. The BP network, a widely used neural network model, can learn and store numerous input-output pattern relationships without requiring mathematical equations. Though individual neurons are simple, a network of them can exhibit diverse behaviors. Following a systematic approach, we used the BP neural network to predict the likelihood of successfully hosting the Olympic Games based on six indices. These indices, which include GDP, national population, service industry value-added, prime age labor force proportion, total national reserves, and medical level, were obtained from the International Bureau of Statistics. The training graph shows that when the value exceeds 1, the Games can proceed as planned; otherwise, hosting is likely to be unsuccessful.

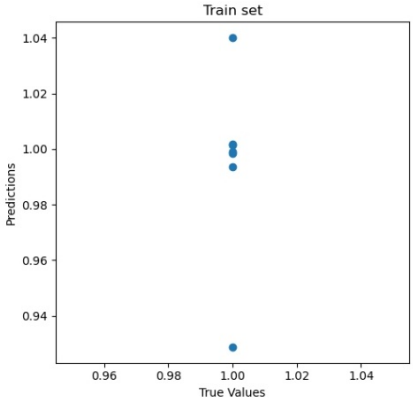

Fig. 3 BP neural network training effect graph

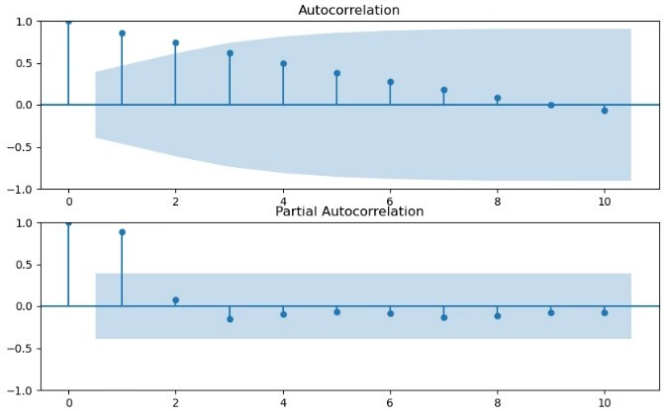

Fig. 4 Autocorrelation and Partial-Autocorrelation graph

We use a Markov process to predict the feasibility of hosting the Olympic Games, in cases where no country shows interest. The method forecasts future states based on current states, using a transition probability matrix. Using the United States as an example, we analyzed eight years of Olympic Games bidding data (2020-2028) to predict its capability to host the 2028 Games. The

prediction confirmed the U.S. meets the requirements, matching the reality of the Games taking place there, proving textual method's viability.

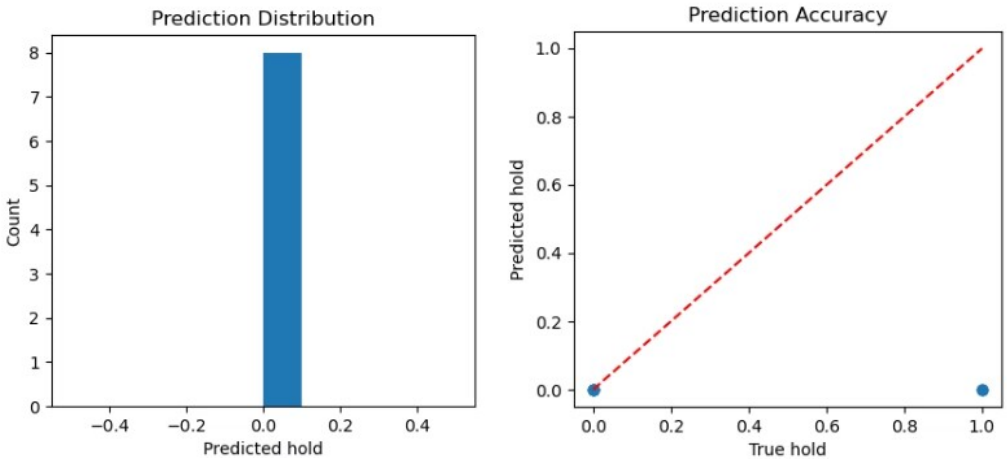

Fig. 5 Prediction distribution and accuracy graph

3. Promotion and Realization Of The Model

Multiple linear regression models are beneficial as they use multiple independent variables for better predictions. Vector Autoregression (VAR) calculates the probability of strategy generation over time and predicts future data movements. The Back Propagation (BP) neural network model can approximate nonlinear functions, making it suitable for complex influence problems. However, each model has its limitations. Regression analysis can be constrained by the speculative selection of a factor. VAR, while useful, is not the perfect solution for managing price risk and assumes independence between consecutive trading days. The BP algorithm can face issues like the "sawtooth phenomenon" and minimal changes in weight errors due to its complex objective function. Lastly, the Markov prediction model can be subjective as it considers the weights of indicator cases as given.

References

[1] MACALOON J J. Agenda 2020 and the Olympic movement[J]. Sport in Society, 2016, 19(6): 767-785. [2]

[2] CHAPPELET J L. The International Olympic Committee and the Olympic system：The governance of world sport[M]. New York: Routledge, 2008: 177-181.

[3] CHAPPELET J L. From Olympic administration to Olympic governance[J]. Sport in Society, 2016, 19(6): 739-751.

[4] Matheson, V. and Zimbalist, A. (2021, April 19). Why Cities No Longer Clamor to Host the Olympic Games. Georgetown Journal of International Affairs. Found February 9, 2023 at: https://gjia.georgetown.edu/2021/04/19/why-cities-no-longer-clamor-to-host-the-olympic-games/.

[5] Wang Yufei, Du Tiancang. Analysis and Application of Holt-Winters-based Time Series Forecasting in Big Data Monitoring System [J] .Computer and Modernization, 2019 (11): 29-33.

[6] Ruan Jie. Research on Fuzzy Evaluation of Effectiveness of Regional Independent Innovation Alliance Construction [J]. Modern Business Industry, 2019, 40 (35): 20.

[7] Prashant Kumar Shrivastava;Arun Kumar Pandey Multi-Objective Optimization of Cutting Parameters during Laser Cutting of Titanium Alloy Sheet using Hybrid approach of Genetic Algorithm and Multiple Regression Analysis [J] Materials Today: Proceedings, 2018.