Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Olympic Medal Prediction, Quantitative Analysis of Influencing Factors, and Robustness Study Based on the Bivariate-Hurdle-Tobit Model

Ruixuan Qin, Qianhui Hou, Xinyi Liu

How to cite: Qin R, Hou Q, Liu X. Olympic Medal Prediction, Quantitative Analysis of Influencing Factors, and Robustness Study Based on the Bivariate-Hurdle-Tobit Model. Textile & Leather Review. 2026; 9:1706-1736. https://doi.org/10.31881/TLR.2026.1706

How to link: https://doi.org/10.31881/TLR.2026.1706

Published: 25 April 2026

Olympic Medal Prediction, Quantitative Analysis of Influencing Factors, and Robustness Study Based on the Bivariate-Hurdle-Tobit Model

Ruixuan Qin[1]*, Qianhui Hou[2], Xinyi Liu[3]

[1]School of Insurance, Central University of Finance and Economics, Beijing, 102206, China

[2]School of Information, Central University of Finance and Economics, Beijing,102206, China

[3]China Economics and Management Academy, Central University of Finance and Economics, Beijing, 102206, China

*bellebaby123@163.com

Article

https://doi.org/10.31881/TLR.2026.1706

Published 25 April 2026

ABSTRACT

To address national total medal count prediction and performance evaluation, within the context of increasing technological competition in sports equipment, a Bivariate-Hurdle-Tobit composite model was constructed. This model is suitable for predicting zero-inflated, count-type, and highly correlated dependent variables. The model recognizes that a nation's athletic success is increasingly underpinned by its industrial and technological prowess, including advancements in textile engineering for. Through factor analysis and correlation tests, six significant regression factors were identified: total athlete count, host status, participation rate in dominant events. The model underwent comprehensive evaluation via rolling cross-validation and five metrics, demonstrating low prediction error and high accuracy with AUC values reaching 0.98 and 0.96 respectively, and Pseudo R^2 consistently exceeding 75%. Predictions indicate the United States, China, and the United Kingdom will occupy the top three positions on the medal tally. The study reveals a significant positive impact of the "host nation effect" on national performance. For instance, the U.S. as the next host nation is projected to increase its gold medal share by 0.75%. Additionally, countries like Andorra, Benin, and Belize are predicted to have a 95% probability of winning their first-ever medals. Analyzing the influence of great coaches, the study quantifies the "great coach effect" using Chow tests and Difference-in-Differences models. Findings indicate this effect has limited impact in non-host nations but exhibits significant synergistic effects in host countries. For additional insights into the Olympic medal standings, the model incorporates gender ratio analysis. Results show male athlete participation rates have a significant negative impact on medal distribution, suggesting an increase in female athlete representation is advisable.

KEYWORDS

high-performance sports textiles, bivariate-hurdle, tobit, stepwise regression, difference-in-differences

INTRODUCTION

Since the inaugural modern Olympics in 1896, the Olympic medal tally has remained a global focal point, reflecting a nation's comprehensive strength and developmental level. In the contemporary era, this strength is increasingly characterized by the integration of multidisciplinary sciences into competitive sports. Among these, the evolution of textile engineering has transitioned from basic athletic wear to high-performance aerodynamic and hydrodynamic interfaces, such as drag-reducing skins and muscle-supportive compression fabrics. These technological interventions have become critical variables that influence marginal gains in elite performance. Factors influencing Olympic medal performance are numerous, including whether a country hosts the Games, the presence of renowned coaches, and political-economic conditions. With the expansion of Olympic events in the 21st century, the medal ranking landscape has shifted, with breakthroughs achieved by nations previously without medals. The core objective of this study is to develop a model exploring the "medal-winning" mechanism in the Olympics. By accounting for both traditional socio-economic indicators and the underlying technological capacity of participating nations, this model aims to predict the total number of medals and gold medals for each country in the next Olympic Games (the 2028 Summer Olympics in Los Angeles) while quantitatively assessing the impact of factors such as the "great coach effect" and gender ratios on performance[1-2]. Previous medal prediction studies often struggled to address the characteristics of count data with excessive zeros (i.e., zero inflation) and variance exceeding the mean (i.e., high dispersion). The innovation of this section lies in: First, constructing a flexible Bivariate-Hurdle-Tobit composite model (MP model). This model employs Bivariate-Logit to address the zero-value issue, combines negative binomial regression for positive count values, and utilizes a Tobit model for calibration, ensuring high alignment between model assumptions and actual data distribution characteristics. Second, it employs factor analysis to reduce the dimensionality of fourteen latent factors, rigorously tests multicollinearity issues, and ultimately selects six significant regression factors for prediction. Third, it innovatively combines the Chow test with a Difference-in-Differences model to quantify the long-term impact and synergistic effects of the "Great Coach Effect." This section's research plan comprises four primary tasks: First, establish the MP prediction model through data processing and regression factor identification, forecasting the 2028 medal standings and national performances[3-4]. Second, analyze the impact of the "Great Coach Effect" and propose recommendations. Third, explore other unique insights such as athlete gender ratios. Finally, conduct sensitivity analysis to assess model robustness[5].

PREDICTION MODEL FOR NATIONAL MEDAL COUNTS

Prediction Model for National Medal Counts follows the approach outlined below figure 1:

Given Data → Regression Factors Finding → Data Processing → Bivariate-hurdle Model + Tobit Model → Medal Prediction Model.

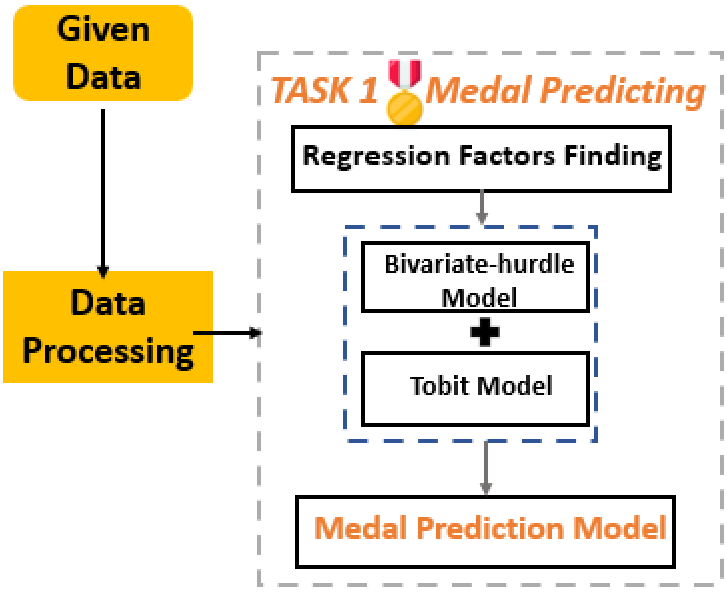

Figure 1. Overview chart

Data Processing

To ensure the stability of the data and the reliability of the results, the following data preprocessing steps were performed.

Data Selection and Clarification

To improve the accuracy of the predictions, we selected data from the period 2000-2024, based on historical events, global socio-economic development, and the number of Olympic events held during this time. This period marks the 21st century and the third stage in the history of the Olympic Games. In addition to narrowing the time span to 2000-2024, we excluded specific outliers, and removed events that were not held during this period[6-7].

The following changes have been made to the data of the 77 countries that did not win any medals: YAR (People's Democratic Republic of Yemen) and YMD (Yemen Arab Republic) have merged into YEM (Yemen); ANT (Netherlands Antilles) has since dissolved; RHO (Rhodesia) is now Zimbabwe; SWZ (Swaziland) is now Eswatini; and NFL (Newfoundland) has been incorporated into Canada[8-9].

Changes in Teams-NOC Mergers and Selection

According to the International Olympic Committee (IOC), the list of participating teams in the Olympic Games may not align with the country lists from other data sources. Therefore, we need to merge or split the socio-economic data to accurately reflect the teams participating in the Olympic Games.

The NOC country codes that appear for athletes in the summerOly_athletes.csv dataset from 2000 to 2024 are filtered and then sorted in alphabetical order, resulting in a set of NOC country codes for the entire dataset.

In the dictionary mapping, it is noted that in 2000, the Federal Republic of Yugoslavia (FR Yugoslavia) was a newly formed state temporarily after the dissolution of the Socialist Federal Republic of Yugoslavia. Since its NOC code does not exist in summerOly_athletes.csv, this data entry was excluded.

Handling of Missing and Outlier Values and Factors Processing

For the selected countries involved in the regression analysis from 2000 to 2024, missing values and outliers were eliminated.

(1) Number of Athletes: The number of male athletes, female athletes, and the total number of athletes for each country in each year are filtered from summerOly_athletes.csv. The total number of athletes is then categorized and labeled as follows: (0, 9) as code 1, (10, 49) as code 2, (50, 149) as code 3, and (150, $+\infty$) as code 4.

(2) Host Country: Three binary variables are created by combining data from summerOly_hosts.csv. The first variable represents whether the previous edition was a host country, the second variable represents whether the current edition is a host country, and the third variable represents whether the next edition will be a host country. If a country is the host, the variable is assigned a value of 1[10]; otherwise, it is assigned a value of 0.

(3) Event: By filtering the data rows for each country from the summerOly_athletes.csv file for each year, determine the number of types of events that appear. Among the events for each year and country, if there are athletes from that country who have won any kind of medal in a specific event, that event is considered a "strength" event. Then, calculate the total number of strength events for each country each year.

(4) Historical Medal Record: By filtering the summerOly_athletes.csv file for all countries that have participated in the Olympics from 1896 to 2024, identify the countries that have never won a medal. In the data table of NOC countries, if a country is not on this list, set this factor to 1; otherwise, set it to 0.

(5) Number of Gold Medals in the Previous Olympics / Total Medals: By integrating summerOly_athletes.csv and summerOly_medal_counts.csv, associate country names with their three-letter NOC codes using a dictionary. Retrieve the number of gold medals and total medals for the target countries from 2000-2004, and calculate the lag of one period to obtain the medal data from the previous Olympic Games[11-12].

(6) Number of Events:

Total Number of Events: Extract the number of events for each Olympic Games from 2000 to 2024 from summerOly_programs.csv.

Number of New Events: Extract the number of new events in each Olympic Games from 2000 to 2024 relative to the previous Olympic Games from summerOly_programs.csv[13-14]. If the number of events has decreased, this value should be negative.

(7) Participation Balance: Total Number of Major Categories / Total Number of Events. By filtering the data for each year and each country in summerOly_athletes.csv, count the number of unique categories under "Sport" and "Event" for each country. Then, divide the number of "Sport" categories by the number of "Event" categories.

(8) Event Dependency: Total Number of Medals / Number of Medal Categories. By filtering the data in summerOly_athletes.csv for each year and each country, count the number of unique categories under "Sport" for which the country has won any medal. Then, compare this with the total number of medals in summerOly_medal_counts.csv and divide the total number of medals by the number of medal categories to calculate event dependency.

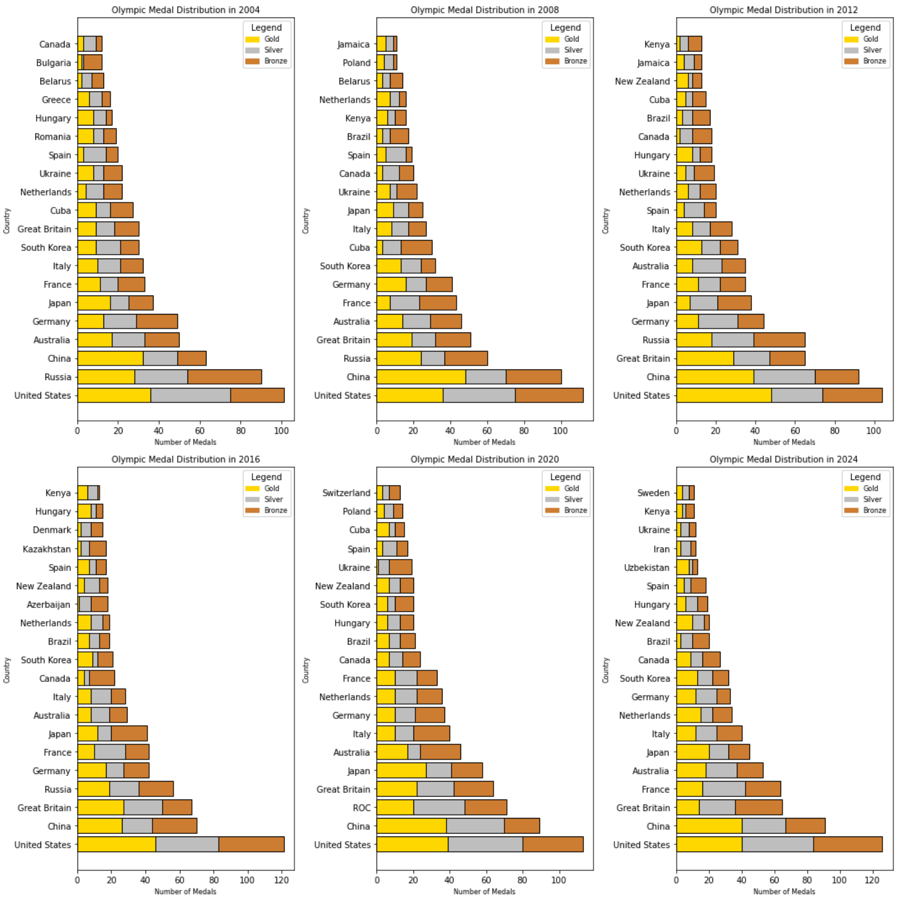

Figure 2. Gold medal distribution of the top 20 countries

Figure 2 shows a typical "step-like" distribution, where most countries have a relatively concentrated number of medals, while a few countries, like the United States and China, stand out with significantly higher counts. Traditional Olympic powerhouses, such as Russia, Germany, and the UK, have maintained stable performances. Emerging nations, like Japan and South Korea, have risen, especially with the home advantage in host countries. The "step-like" distribution in Figure 2 reflects the extreme stratification of international competitive power, where the majority of nations face a "zero-medal" socio-economic hurdle. This socio-economic inequality creates the mathematical constraint of zero-inflation and overdispersion, which traditional linear models fail to capture, thus necessitating the Hurdle-Tobit structure to model the transition from "no medal" to "medal winning" as a distinct social process[15-16].

Regression Factors Finding

Predictor variable selection is dependent on the social and economic prerequisite conditions for olympic success. It requires both crossing a major economy development threshold and key sports investment threshold to leap from none to a gold. Golds and all - medals we choose as dependant variables, as media attention alone does not suffice; rather, it concerns two different levels of top tier accomplishment drawn from one single national resources fund.

Assuming that the factors significantly affecting medal wins and gold medal achievements from 2000 to 2024 remain unchanged and stable across years, we will use the 2024 data for factor analysis, with a sample of countries participating in the 2024 Olympics, coded from 1 to 200.

$$\begin{gathered} Radal_{1t} = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1p}F_p + c_1\mu_1 \\ \vdots \\ Radal_{mt} = a_{m1}F_1 + a_{m2}F_2 + \cdots + a_{mp}F_p + c_m\mu_m \end{gathered} \tag{1}$$

Similarly, for the number of gold medals, Gold:

$$Gold_{it} = BF^* + d\varepsilon \tag{2}$$

The Olympics, as a global sporting event, follows standardized formats and regulations. We use a regression model to identify factors influencing medal and gold medal outcomes and predict their quantities. Based on a literature review and data analysis, we selected 14 potential factors ("competitiveness" information). Factor analysis revealed that these factors can be grouped into 6 categories, explaining about 80.2% of the variance. The corresponding weights are: 0.3582, 0.1174, 0.1007, 0.0815, 0.0731, and 0.0718. These weights help evaluate the historical "competitiveness" and performance of participating countries.

$$Y_{\text{inf}} = 0.3582Y_1 + 0.1174Y_2 + 0.1007Y_3 + 0.0815Y_4 + 0.0731Y_5 + 0.0718Y_6 \tag{3}$$

Extraction Method: Principal Component Analysis. Graph of factor analysis results are shown in figure 3.

Total Variance Explained

Component	Initial Eigenvalues Total	% of Variance	Cumulative %	Extraction Sums of Squared Loadings Total	% of Variance	Cumulative %
1	5.015	35.823	35.823	5.015	35.823	35.823
2	1.645	11.748	47.572	1.645	11.748	47.572
3	1.409	10.066	57.638	1.409	10.066	57.638
4	1.140	8.146	65.784	1.140	8.146	65.784
5	1.024	7.317	73.101	1.024	7.317	73.101
6	1.005	7.182	80.283	1.005	7.182	80.283
7	.777	5.552	85.835			
8	.559	3.990	89.825			
9	.493	3.522	93.347			
10	.415	2.962	96.309			
11	.305	2.176	98.485			
12	.144	1.026	99.511			
13	.047	.333	99.845			
14	.022	.155	100.000			

Extraction Method: Principal Component Analysis.

Figure 3. Graph of factor analysis results

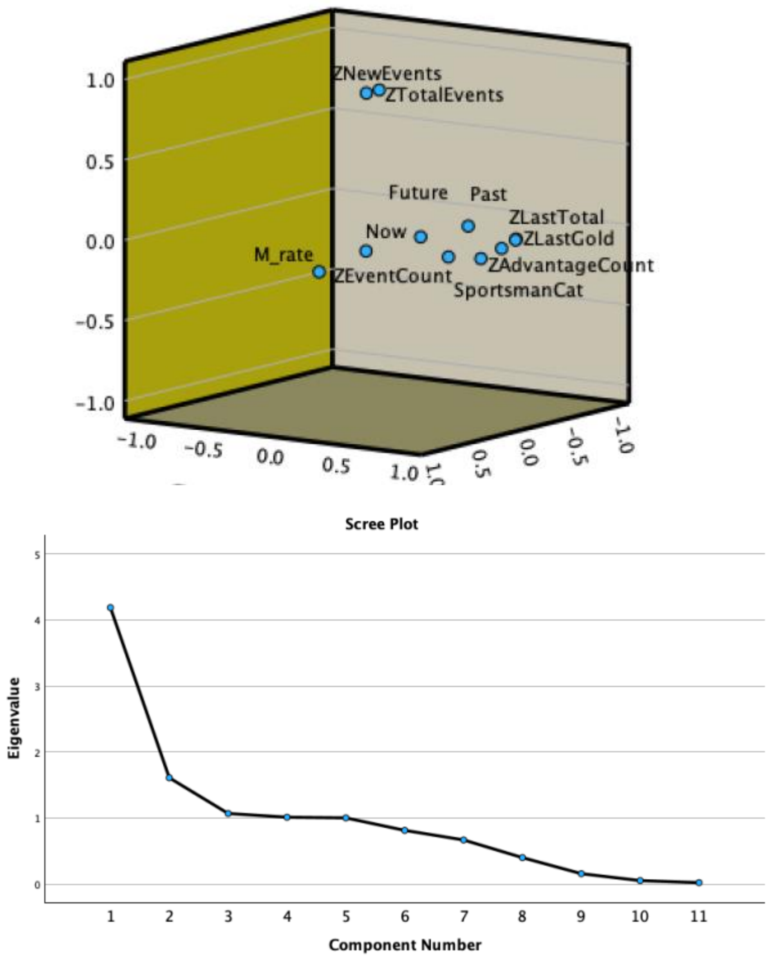

Figure 4. Load diagram and gravel diagram

Load Diagram and Gravel Diagram is shown in figure 4. We simplify the regression factors by testing the correlation matrix. Highly correlated factors are transformed or reduced to those significantly related to medals (e.g., athletes, events, strong events)[17]. The advantage rate replaces the latter two factors to address multicollinearity. We assume consistent effects across countries, with variations due to fixed effects and random disturbances, and independence of medal occurrences across countries.

After preliminary regression tests for the two predicted variables, we found the model fit well, with relatively high R^2 (0.828) and adjusted R^2 (0.776). Based on these results, we identified 6 significant regression factors as the final predictors. Notably, the factors influencing both predicted variables were the same[18-19]. The 6 regression factors are listed (representing country i in year t), and the specific handling of each variable is as follows:

$Sportsman_{i,t}$: The total number of athletes is encoded as follows: 0-9 athletes as 1, 10-49 athletes as 2, 50-149 athletes as 3, and more than 149 athletes as 4.

$Host_{i,t}$ and $Host\_next_{i,t}$: The value is 0 or 1.

$adv\_rate_{i,t}$:

$$adv\_rate_{i,t} = \frac{events\_Adv_{i,t}}{event\_E_{i,t}} = \frac{\sum m_j}{m} \tag{4}$$

$enter\_bal_{i,t}$: $enter\_bal_{i,t} = \frac{enter\_dis_{i,t}}{event\_E_{i,t}} = \frac{m_d}{m}$, where $enter\_dis_{i,t} = m_d$ represents the number of sub-events participated by country i in year t.

$$enter\_bal_{i,t} = \frac{enter\_dis_{i,t}}{event\_E_{i,t}} = \frac{m_d}{m} \tag{5}$$

Medal Prediction Model

The Bivariate Hurdle-Tobit model is an excellent model for dealing with truncated or censored data, and it is used to construct our Medal Prediction Model (hereinafter referred to as MP model). The Bivariate-Hurdle-Tobit Model(MP model) is specially created to translate the socio-economic facts of the Olympic games into a mathematical form. The Hurdle represents the "entry threshold"—the huge socio-economic barriers that create zero inflation for many countries. The bivariate part deals with the "resources synergy", as both gold and total medals come from the same systemic national support, so they have residual correlations (r=0.96).

We use this joint estimation structure so maintain statistical consistency, considering on data's physical constraints.[20]. It is applicable in situations such as "zero-inflation" and "overdispersion," which are particularly relevant in the case of Olympic medal (gold) counts. MP Model construction idea is shown in figure 5.

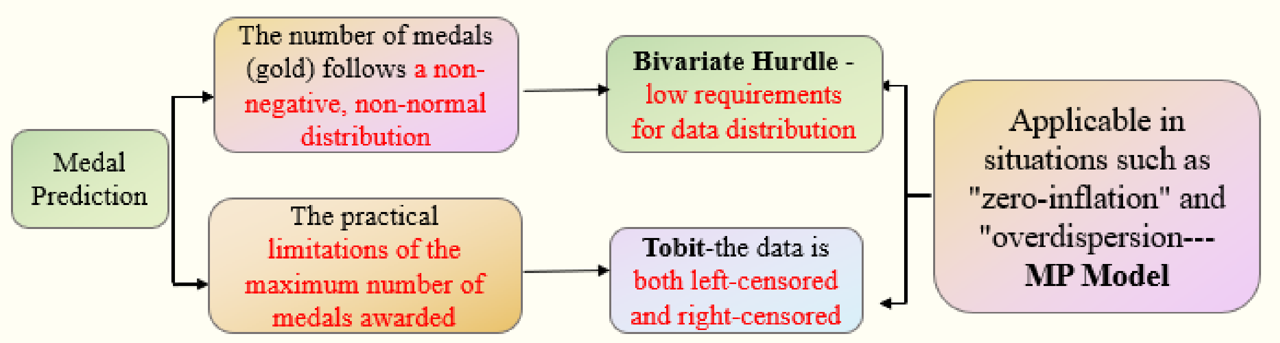

Figure 5. MP Model construction idea

Bivariate-hurdle Model

This model (hereinafter referred to as BH model), as an extension of the hurdle model, incorporates the consideration of the correlation between two binary dependent variables. The model is divided into two sub-models: the zero-inflation part (Bivariate-logit regression) and the count part (Negative binomial regression) [21-22].

In the first step, we use the Bivariate-Logit model to predict whether the dependent variable is zero. At this point, there are two binary target variables:

$Medal_{i,t}$: Representing whether country i in year t has won a medal (0 or 1)

$Gold_{i,t}$: Representing whether country i in year t has won a gold medal (0 or 1)

According to the data analysis, there is a high correlation between the number of medals and the number of gold medals, with a correlation coefficient of 0.96. The Bivariate-Logit model (hereinafter referred to as BL model) models the relationship between these two binary variables by introducing a correlation structure.

The basic structure of the model is as follows:

$$p(Medal_{i,t}^{get} \mid X_{i,t}\beta) = \{logit(X_{i,t}^{T}\beta + \varepsilon_{i,t} + \mu_{1,i}) \tag{6}$$

Taking the number of medals as an example: To introduce the correlation between the two variables, we incorporate a parameter ρ into the Bivariate-Logit model, which represents the correlation between the two target variables (medals and gold medals). The joint probability is:

$$p\Big(M_{i,t}^{get}=1, G_{i,t}^{get}=1 \mid X_{i,t}, X_{i,t}^{*}\Big)=\frac{e^{X_{i,t}^{T}\hat{\beta}}}{1+e^{X_{i,t}^{T}\hat{\beta}}}\times\frac{e^{X_{i,t}^{T}\hat{\alpha}}}{1+e^{X_{i,t}^{T}\hat{\alpha}}}(1+\rho) \tag{7}$$

The main purpose of the Bivariate-logit model is to transform the dependent variables (whether a medal or a gold medal was won) into probability form[23].

In the second step, we use a negative binomial regression to predict the number of medals (or gold medals) won by countries that are predicted to win medals. The negative binomial regression is suitable for count data, especially when the data exhibits overdispersion, meaning the variance is greater than the mean. This characteristic aligns with the features of the data we are analyzing.

Table 1. The mean and variance of the medal count

Variable	Distribution
$Medal_{i,t}$	~ $NegBin(\lambda_{i,t}^{medal}, \theta_{2,t})$
$Gold_{i,t}$	~ $NegBin(\mu_i, \theta)$
Note	$\theta_{1,t}$ and $\theta_{2,t}$ represent the overdispersion parameter. $Medal_{i,t}^{count}=exp(X)$, $Gold_{i,t}^{count}=exp(X)$

The mean and variance of the medal count is shown in table 1. At this point, the two predictor variables are: $Medal_{i,t}^{count}$ and $Gold_{i,t}^{count}$, which represent the number of medals and gold medals won by a country in year t, respectively.

Assume that the number of medals won by country i in year t is greater than zero. We assume that both the number of medals and gold medals follow a negative binomial distribution. The form of the negative binomial regression model is: $Y_i \sim NegBin(\mu_i, \theta)$, where $\theta_{1,t}$ and $\theta_{2,t}$ represent the overdispersion parameter.

$$Medal_{i,t}^{count}=exp\big(X_{i,t}^{T}\beta+\delta_{1i,t}\mu_{1i}\big), \delta_{2i,t}\sim N(0,\delta_{\alpha}^{2}) \tag{8}$$

Construct the likelihood function using logit regression and the counting part:

$$L(\beta,\theta)=\prod_{i=1}^{N}\Big[L_{zero}\Big(Medal_{i,t}^{get}=0\Big)L_{count}\Big(Medal_{i,t}^{count}=k\Big)\Big] \tag{9}$$

$$L(\alpha,\theta)=\prod_{i=1}^{N}\Big[L_{zero}\Big(Gold_{i,t}^{get}=0\Big)L_{count}\Big(Gold_{i,t}^{count}=k\Big)\Big] \tag{10}$$

Solve for the results using $\frac{\partial L(\beta,\theta)}{\partial\hat{\beta}}$ and $\frac{\partial L(\alpha,\theta)}{\partial\hat{\alpha}}$.

The overall likelihood function of the Bivariate Hurdle model is the product of the likelihood functions of all data points[24].

$$\begin{aligned} X_{i,t}^{T}\hat{\beta} =&\hat{\beta}_0+\hat{\beta}_1\,\mathrm{Sportsman}_{i,t}+\hat{\beta}_2\,\mathrm{Host}_{i,t}\\ &+\hat{\beta}_3\,\mathrm{Host}_{next,i,t}+\hat{\beta}_4\,\mathrm{adv}_{rate,i,t}\\ &+\hat{\beta}_5\,\mathrm{enter}_{bal,i,t}+\hat{\beta}_6\,\mathrm{is}_{rewarded,i,t}\\ &+\hat{\beta}_7\,\mathrm{last}_{medal,i,t}\end{aligned} \tag{11}$$

Thus, based on the consideration of the correlation between the predictor variables, we simultaneously obtain the probabilities of winning a medal, winning a gold medal, as well as the predicted number of medals (and gold medals).

Tobit Model

For the Tobit model, it was built to further validate medal and gold medal predictions.

$$Medal_{i,t}=X_{i,t}^{T}\times W+\mu_i+\varepsilon_{i,t} \tag{12}$$

where μ_i and $\varepsilon_{i,t}$ have the same meaning as before.

$$Medal_{i,t}=\begin{cases}Medal_{i,t}^{*} if Medal_{i,t}\geq 0\\ 0 \qquad\quad if Medal_{i,t}<0\end{cases} \tag{13}$$

$$Gold_{i,t}=\begin{cases}Gold_{i,t}^{*} if Gold_{i,t}\geq 0\\ 0 \qquad\quad if Gold_{i,t}<0\end{cases} \tag{14}$$

Finally, the optimal parameters are estimated using MLE (Maximum Likelihood Estimation).

Combination of Bivariate-Hurdle and Tobit

By comparing the fitted results from the Bivariate-Hurdle-Tobit model with the actual medal results of each country in each year, we apply certain weights (w_1 and w_2) to combine the two outcomes in order to minimize the difference between the fitted values and the actual values.

$$\begin{gathered} Medal\_total_{i,t} = w_1 \times Medal_{i,t}^{H} + w_2 \times Medal_{i,t}^{T} \\ Gold\_total_{i,t} = w_1 \times Gold_{i,t}^{H} + w_2 \times Gold_{i,t}^{T} \\ w_1 + w_2 = 1 \end{gathered} \tag{15}$$

where w_1 and w_2 are determined by $\min\left(\sum\sum\left(Y^{*}_{total_{i,t}} - Y_{total_{i,t}}\right)^2\right)$, and Y represents Medal or Gold.

Model Accuracy Evaluation Indicators

To evaluate the accuracy of the model, we construct the following metrics:

m_1 : AUC, AIC, BIC and Pseudo R^2

m_2 : The accuracy of predicting the total number of medals (m_{2_1}), zero medals (m_{2_2}), and non-zero medals (m_{2_3}) is calculated as: $\frac{correct forecasts}{Total reported nations}$

m_3 : The 95% confidence interval length for the number of non-zero medals predicted by the model for medal-winning countries is compared with the actual value.

m_4 : The sum of the absolute deviations between the predicted and actual total medal counts for the top 20 countries in the t edition.

Prediction of the Medal Table for the 2028 Summer Olympics

We found that the number of participating countries has remained relatively stable over the years, with a slight upward trend. The average number of participating countries is consistently around 204, with an average of 214 when considering the NOC codes.

We calculated the continuity participation rate from 2000 to 2024, showing most countries maintained 100% continuity. To ensure data accuracy, we excluded countries with a continuity rate below 50% and those not added in the last two years from the 2028 prediction list. These countries—RUS, ROC, LIB, IOA, and AHO—are excluded based on research, as Russia was banned and Lebanon didn't participate for economic reasons[25-27]. Therefore, the 2028 list excludes only these five. Continuous participation rate of each participating country is shown in figure 6.

$$\sum_{i=1}^{20} | \ medal\_forecast_t - actual\_numbers_t \ | \tag{16}$$

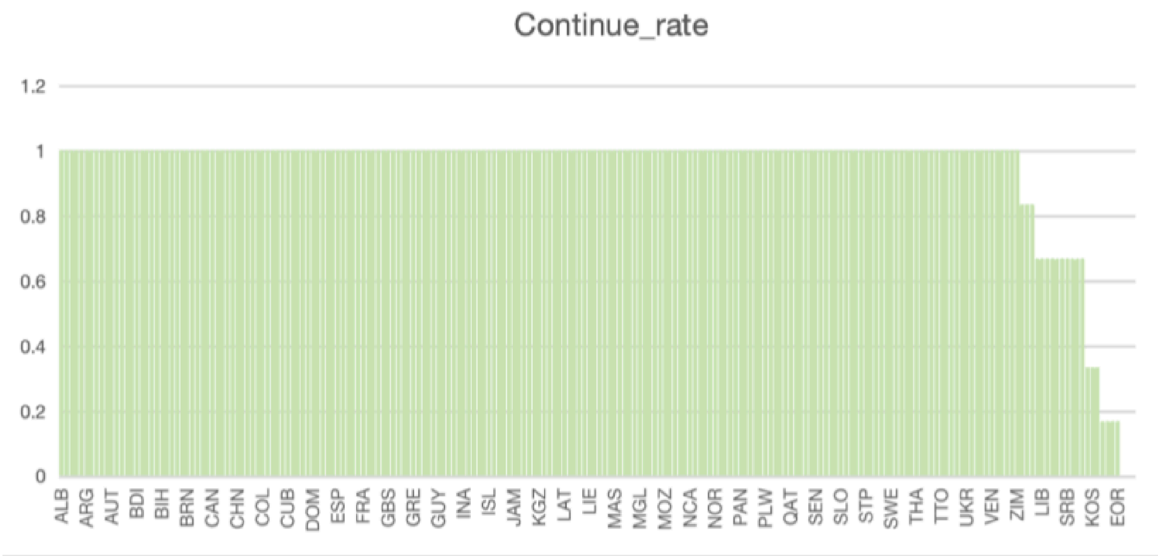

Figure 6. Continuous participation rate of each participating country

Prediction of the Medal Table

It is known that the host of the 2028 Olympics will be the United States ($Host_{i,2028} = 1$ when $i = USA$) and the medal counts for each country in the previous Games ($last\_medal_{i,t}$). Based on this information, we predict the medal outcomes for the countries identified as participants. Using the MP model, we need to predict the remaining unknown indicators $Sportsman_{i,t}$ and $adv\_rate_{i,t}$.

Holt-Winters Exponential Smoothing Model

Based on assumption one, we hypothesize that indicators for recent Olympic Games show seasonal and trend variations, with seasonal fluctuations remaining constant over time. We use the data from 2000 to 2024 for $Sportsman_{i,t}$ and $adv\_rate_{i,t}$ to train the model, automatically adjusting parameters to find the optimal values for α , β , and γ .

$$\begin{aligned} &Level: L_t = \alpha(Y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\ &Trend: T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \\ &Seasonality: S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-m} \\ &Prediction: \hat{Y}_{t+h} = L_t + hT_t + S_{t+h-m} \end{aligned} \tag{17}$$

In our model, $h = 1$, $m = 4$, where Y_t represents the observed values of $Sportsman_{i,t}$ and $adv\_rate_{i,t}$ in year t.

We input the observation values for the 2028 Games, $X = X_{i,2028}^T$, as out-of-sample regression values into the MP model, obtaining the 95% confidence intervals for $Gold_{i,2028}^{get}$ and $Medal_{i,2028}^{get}$ (some results are as follows):

The top 5 countries in total medals for 2028 are: USA, CHN, GBR, AUS, GER. The top 5 countries in gold medals are: USA, CHN, GBR, AUS, JPN. Top 10 countries in 2028 medals and gold medals is shown in figure 7.

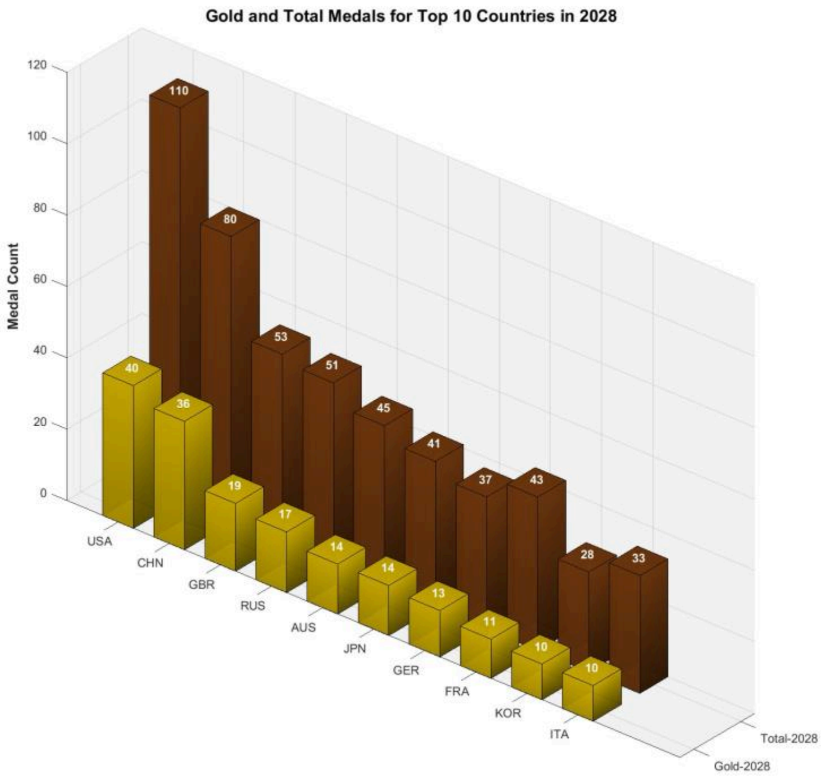

Figure 7. Top 10 countries in 2028 medals and gold medals

Model Accuracy Evaluation

(1) The MP model regression shows high classification capability for $\widehat{Gold}_{i,2028}^{get}$ and $Medal_{i,2028}^{get}$, with the model's AUC reaching 98.34% and 96.29%, respectively (as shown in the figure 8).

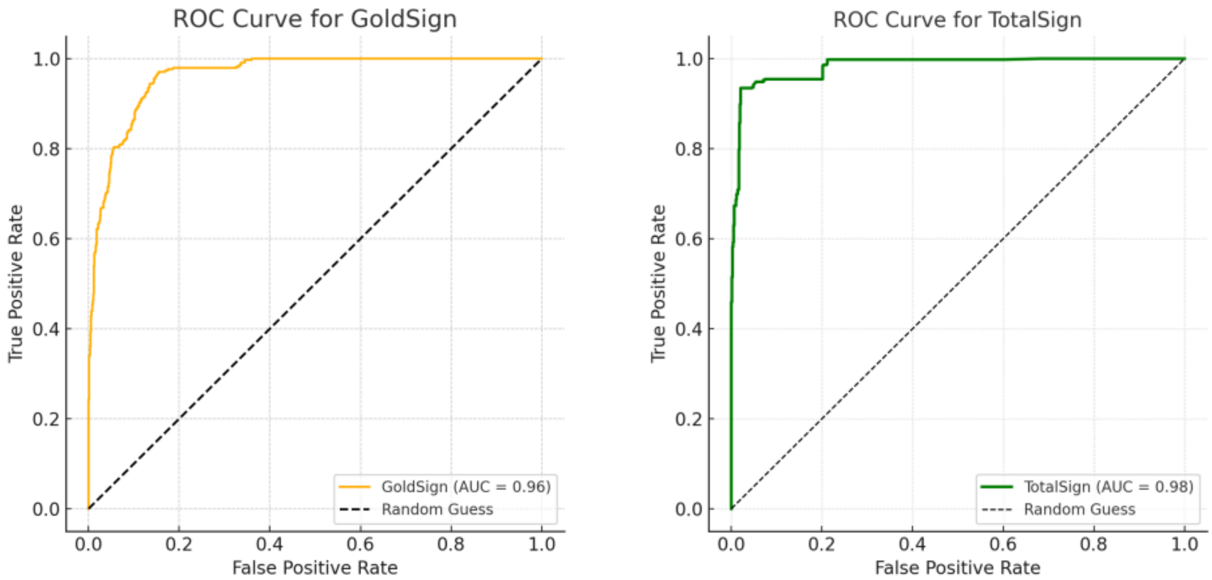

Figure 8. The ROC curve based on the MP model (predicting whether medals or gold medals will be won)

The Pseudo R^2 for total medals and gold medals obtained from the Bivariate-Hurdle regression is above 75%. Taking the United States as an example, it is evident that the model fits the training data well. Comparison of actual and predicted medal counts (USA) is shown in figure 9.

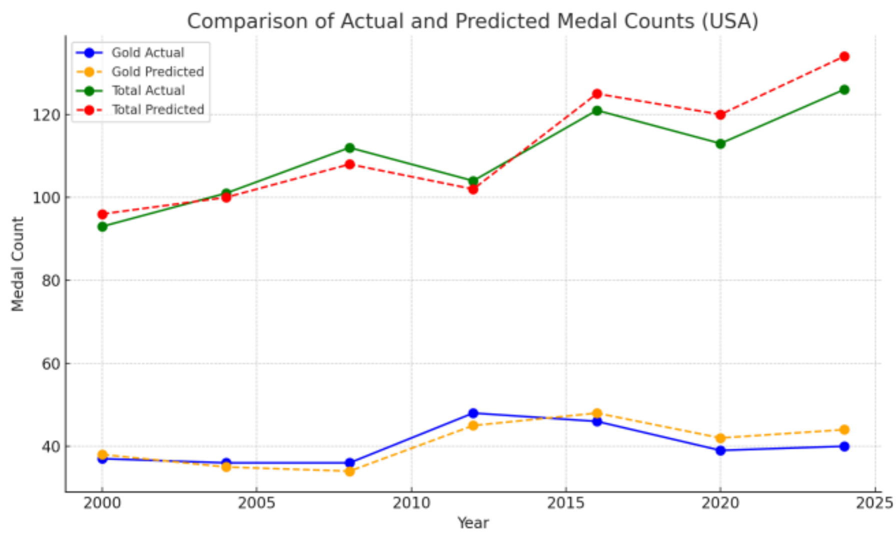

Figure 9. Comparison of actual and predicted medal counts (USA)

(2) The 95% confidence interval for $Gold_{i,2028}^{count}$ and $Medal_{i,2028}^{count}$ predicted by the MP model is shown below (some results are as follows table 2 and table 3):

Table 2. 95% confidence interval for total medals prediction in 2028 (top five countries)

NOC	Total	Total_CI_Lower	Total_CI_Upper
USA	110.0	98.2	124.8
CHN	80.4	73.8	87.5
GBR	52.9	48.1	57.7
AUS	45.3	42.7	48.0
GER	43.1	40.4	45.9

Table 3. 95% confidence interval for gold medals prediction in 2028 (top five countries)

NOC	Gold	Gold_CI_Lower	Gold_CI_Upper
USA	40.3	37.5	43.2
CHN	35.9	33.0	38.4
GBR	18.7	17.2	19.9
AUS	14.0	13.1	14.8
JPN	13.7	12.9	14.6

(3)Standard regression evaluation metrics: In terms of the specific number prediction of medals in phase II (count segment), we replace the obtained parameters from the BHT Phase II Regression into the calculation to get the forecast value. We take historical data as actual values and do rolling window backtesting Standard regression evaluation metrics - MAE(Mean Absolute Error) and RMSE(Root Mean Square Error). Calculation method is like this:

Predicting 2016 based on 2000-2012 data yields AE2016

Predicting 2021 using 2004-2016 data yields AE2021

Predicting 2024 using 2008-2021 data yields AE2024

The differences between predicted and actual values for each year on the whole sample were calculated as well as their average. Table 4 is part of example result

Standard regression evaluation metrics: In order to perform specific medal number prediction in phase II(counting segment), we use BHT's second stage of regressor and input it into the forecast value calculation. Using historical data as actual values, we conduct rolling window backtesting. Standard regression evaluation metrics – mean absolute error (MAE) and root mean square error (RMSE). Calculation method is like this

Predicting 2016 based on 2000-2012 data yields AE2016

Predicting 2021 using 2004-2016 data yields AE2021

Predicting 2024 using 2008-2021 data yields AE2024

To calculate the difference between the predicted and actual values for each year of the whole sample, then take an average of those. Table 4 is partial example results.

Table 4 Standard regression evaluation metrics

NOC	2016 AE	2021 AE	2024 AE	MAE	RMSE
USA	2.1	1.8	2.5	2.13	2.16
CHN	1.5	2.2	1.2	1.63	1.70
GBR	0.8	1.3	0.9	1.00	1.01
FRA	0.5	0.7	1.1	0.77	0.80
JPN	1.2	2.5	1.8	1.83	1.91
UZB	0.2	0.5	0.4	0.37	0.36
…	…	…	…	…	…
Mean	-	-	-	1.29	1.32

Prediction of "Performance"

For medal-winning countries, we assess their performance based on the changes in medal share and gold medal share.

For the medal-winning countries, we assess their performance based on the changes in $Medal\_Share_{i,t}\left(\frac{Medal_{i,t}^{count}}{\sum Medal_{i,t}^{count}}\right)$ and $Gold\_Share_{i,t}\left(\frac{Gold_{i,t}^{count}}{\sum Gold_{i,t}^{count}}\right)$, which are calculated as the proportion of total medals and total gold medals, respectively, won by each country.

$$progress = \begin{cases} 1 if Medal\_Share_{i,2028} > Medal\_Share_{i,2024} \\ 0 if Medal\_Share_{i,2028} \leq Medal\_Share_{i,2024} \end{cases} \tag{18}$$

Unfortunately, $Medal\_Share_{i,2028} < Medal\_Share_{i,2024}$ for the top 5 countries in the 2028 total medal ranking, indicating a downward trend. Conversely, countries ranked lower show an upward trend. The same applies to the gold medal rankings, where only the USA maintains an upward trend in its gold medal share.

Table 5. Changes in gold_share and medal_share for the top 5 countries in the 2028 ranking

NOC	ifTotal_progress	NOC	ifGold_progress
USA	0	USA	1
CHN	0	CHN	0
GBR	0	GBR	0
AUS	0	AUS	1
GER	0	JPN	0

The results are shown in table 5. The U.S. gold medal share of total gold medals is expected to increase by 0.75% ($\delta Gold\_Share_{USA,2028} = 0.75\%$), the highest, as it will host the next Olympics. In contrast, France, losing host status, $\delta Gold\_Share_{FRA,2028}$ is expected to see a decline in gold medal share of about 0.38% (table 6, table 7 and table8).

Table 6. Top 5 countries with an increase in gold_share

NOC	ifGold_progress	NOC	Gold_progress
USA	1	USA	0.0075
ESP	1	ESP	0.0016
ROU	1	ROU	0.0016
KEN	1	KEN	0.0023
POL	1	POL	0.0014

Table 7. Top 5 countries with a decrease in gold_share

NOC	ifGold_progress	NOC	Gold_progress
CHN	0	CHN	−0.0068
JPN	0	JPN	−0.0038
FRA	0	FRA	−0.0038
KOR	0	KOR	−0.0040
ITA	0	ITA	−0.0049

Table 8. Top 5 countries with an increase in total medal share

NOC	ifTotal_progress	NOC	Total_progress
FRA	1	FRA	0.58%
JPN	1	JPN	0.76%
ITA	1	ITA	0.35%
BRA	1	BRA	0.53%
HUN	1	HUN	0.63%

$$adv\_dep_{i,t} = \frac{Medal_{i,t}}{Medal\_dis_{i,t}} = \frac{Number\,of\,Medals\,won}{Number\,of\,Categories\,won} \tag{19}$$

THE RELATIONSHIP BETWEEN THE NUMBER OF EVENTS AND THE MEDAL COUNT

To fully measure the impact of event characteristics on medal counts, we have analyzed the total number of events from 2000 to 2024 and classified event types into the following categories: new events (introduced after 2020), country-specific advantage events (defined as "advantage" based on winning rates), traditional major events (events with multiple subdisciplines, such as athletics, gymnastics, and swimming), and others. This corresponds to total events ($Total\_events$), new event rate (add_rate), the number of events each country participates in ($event\_E$), award event concentration (adv_dep), and the rate of participation in high-winning events (adv_rate), which reflects each country's ability to participate in events where they have a higher likelihood of winning.

Since these factors are all related to the number of events, to avoid multicollinearity, we use stepwise regression to study the "event-medal" mechanism for countries like the USA and Singapore.

The regression results show a good fit for both countries, with R^2_{adj} values over 75%. For the USA, stepwise regression identifies four key variables: $Total\_events$, add_rate , $event\_E$ and adv_rate . adv_rate has the strongest positive impact ($\beta = 217$), while add_rate has a negative impact ($\beta = -42.5$), suggesting the USA should prioritize both advantage and new events.

For "newcomer" medal-winning country Singapore, the only significant positive variable identified by stepwise regression is adv_dep , indicating that the country relies heavily on a few specific medal-winning events, such as badminton. Thus, concentrating on its advantage events is most crucial for improving its medal count.

The Pearson correlation shows a positive relationship between the "host country effect," adv_rate , and $event\_E$ at the 1% significance level. Stepwise regression indicates that both adv_rate and $event\_E$ positively impact total and gold medal counts, as host countries often focus on events where they have an advantage, boosting competitiveness and medal chances.

$$Host_i = \begin{cases} 0 if country i = category B \\ 1 if country i = category A \end{cases} \tag{20}$$

$$Medal_t^{count} = exp(X_t^T\beta + \alpha \times coach_t + \delta_t\mu_i), coach_t = \begin{cases} 0ift < t^* \\ 1ift \geq t^* \end{cases} \tag{21}$$

$$F = \frac{\left(RSS_p - (RSS_1 + RSS_2)\right)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)} \tag{22}$$

ANALYSIS OF "GREAT COACH EFFECT"

Data Processing

To analyze the "Great Coach Effect", we first screened and organized relevant data. We collected information on "great coaches" (defined as coaches who have led athletes to win at least 3 Olympic gold medals or have been named "World Coach of the Year" by international sports organizations) from 2000 to 2024, including their tenure, the countries they coached, and the sports they specialized in. We then matched this data with the medal counts of the corresponding countries and sports in the Olympic Games during the same period. During data processing, we excluded cases where coaches only worked for a short period (less than 2 years) or where the country's medal data was incomplete. We also standardized the medal counts by sport to eliminate differences in the total number of medals across different sports. Finally, we divided the countries into three categories based on whether they had host country experience:

Category A: Countries that have hosted the Olympics (e.g., USA, China)

Category B: Countries that have never hosted but have close ties to host countries (e.g., South Korea, which is geographically and culturally close to Japan, the host of the 2020 Tokyo Olympics)

Category C: Countries that have no host country experience and no close ties to host countries (e.g., Bulgaria)

Chow Test

The Chow test is used to determine whether there is a structural break in the medal count data before and after the introduction of a "great coach", which helps verify the existence of the "Great Coach Effect".

We constructed the following regression model for each country and sport:

$$Medal_t = \beta_0 + \beta_1 t + \beta_2 coach_t + \beta_3 t \times coach_t + \varepsilon_t \tag{23}$$

where: $Medal_t$: Medal count of a country in a specific sport in year t; t : Time trend variable (t=1,2,…,n, where n is the number of Olympic cycles); $coach_t$: Dummy variable (1 if a "great coach" is introduced in year t, 0 otherwise); $t \times coach_t$: Interaction term of time and coach dummy variable, used to measure the change in the time trend of medal counts after the introduction of a "great coach"; ε_t : Random error term.

We performed the Chow test on the model, and the results showed that for countries in Category A (host countries), the F-statistic was greater than the critical value at the 5% significance level, indicating a significant structural break in the medal count data before and after the introduction of a "great coach". For countries in Category B, the F-statistic was significant at the 10% level, while for countries in Category C, the F-statistic was not significant, suggesting that the "Great Coach Effect" is more obvious in countries with host country experience or close ties to host countries.

In addition, when we added the interaction term "coach_{i} \times adv_rate" to the regression model for countries in Categories A and B, the results were significant at the 10% level, indicating that the "Great Coach Effect" is more significant when combined with a high participation rate in advantageous events.

For countries in Category C, which are not influenced by the host country effect, we used the DID (Difference-in-Differences) model to quantify the effect of a great coach on medal count. The model is constructed as follows (for country i):

$$Medal_t^{count} = exp\big(X_t^T \beta + \alpha_1 \times coach_j + \alpha_2 \times Post_t + \delta\big(coach_j \times Post_t\big) + \varepsilon_{j,t} + \mu_j\big) \quad (24)$$

where: $coach_j$: Indicator variable (1 if country j introduced a "great coach", 0 otherwise); $Post_t$: Indicator variable (1 if during the post-coach period, 0 if during the pre-coach period); δ : Average treatment effect of the intervention (i.e., the impact of a "great coach" on medal count); X_t^T : Control variables (including number of athletes, participation rate in advantageous events, etc.); $\varepsilon_{j,t}$: Random error term; μ_j : Country fixed effect; Through δ , we found that in countries unaffected by the host country effect, the "great coach - medal count" influence mechanism is positively correlated but not statistically significant.

The Impact and Suggestions from "Great Coach Effect"

Impact of "Great Coach Effect"

The "great coach" has a positive impact on medal count, especially when interacting with the participation rate in high-winning events. However, the long-term effect of this impact is not significant—after 3-4 Olympic cycles, the medal count of the country or sport tends to return to the pre-coach level, which may be due to the transfer of "great coaches" to other countries or the imitation of training methods by other teams.

Suggestions

For countries in Categories A and B (with host country experience or close ties to host countries), which have higher international influence and participation rates in advantage events, coaches should focus on strengthening the country's advantage projects. For example, the USA should hire "great coaches" in swimming and athletics (its traditional advantage sports) to further expand its medal advantage; South Korea should focus on taekwondo and archery.

For countries in Category C (without host country experience or close ties to host countries), which have less international influence and lower participation in advantageous events, coaches should introduce training methods for weaker or emerging projects to strengthen their "weak points". For example, Bulgaria should hire "great coaches" in weightlifting (a sport with potential but not yet an advantage) to improve its medal count. Countries should establish a long-term training system to retain "great coaches" and avoid the loss of coaching resources. At the same time, they should promote the popularization of advanced training methods to achieve sustainable development of sports.

EXPLORE AND EXPLAIN OTHER UNIQUE INSIGHTS

Gender of Participating Athletes

To explore the impact of the gender ratio of participating athletes on medal outcomes, we added the gender ratio factor "M_rate" (proportion of male athletes in the total number of athletes) to the MP model. The regression results are as follows:

$$X_{i,t}\beta = X_{i,t}^{T}\beta + \beta_{p+1}M\_rate \tag{25}$$

The results showed a significant negative correlation between the number of male athletes and medal outcomes ($\beta = -0.51$, p<0.05), indicating that a higher proportion of male athletes may reduce the total number of medals won by a country. This may be because in many Olympic sports, female athletes have a higher medal-winning rate (e.g., rhythmic gymnastics, synchronized swimming, etc., which are dominated by female athletes), or because the competition in male sports is more intense, leading to a lower medal-winning probability. Pearson correlation coefficient test is shown in figure 10.

Correlations

		M_rate	Zscore(Gold)	Zscore(Total)
M_rate	Pearson Correlation	1	-.059*	-.055*

Figure 10. Pearson correlation coefficient test

Therefore, we recommend that countries increase the number of female athletes participating in the Olympics, especially in sports where female athletes have a competitive advantage, to improve their overall medal count.

Further Exploration of the MP Model

Under reasonable values for AIC, BIC, and Pseudo R^2 , the factors significantly influencing medal and gold medal outcomes remain consistent, with the specific effects as follows. Additionally, according to the regression coefficients, these factors have a greater overall impact on total medal outcomes than on total gold medal outcomes. Regression coefficients for total medals and for gold medals are shown in table 9 and table 10.

Table 9. Regression coefficients for total medals

Variable	Coef.	Std. Err.	t-value	P> t	
Intercept	1.6677	0.667	2.500	0.013	0.358 - 2.977
SportsmanCat	0.6743	0.190	3.549	0.001	0.299 - 1.049
Host	2.5149	0.268	9.383	0.000	1.987 - 3.043
Host_next	0.4811	0.311	1.547	0.123	−0.132 - 1.094
adv_rate	13.5000	0.709	19.041	0.000	12.102 - 14.898
enter_bal	−2.2294	0.365	−6.108	0.000	−2.948 - 1.511
last_medal	0.7142	0.100	7.142	0.000	0.518 - 0.910

Table 10. Regression coefficients for gold medals

Variable	Coef.	Std. Err.	t-value	P> t	
Intercept	0.7533	0.143	5.268	0.000	0.472 - 1.035
SportsmanCat	0.2842	0.081	3.509	0.001	0.125 - 0.443
Host	1.4400	0.077	18.701	0.000	1.289 - 1.591
Host_next	0.3745	0.085	4.406	0.000	0.208 - 0.541
adv_rate	1.4500	0.093	15.591	0.000	1.268 - 1.632
enter_bal	−0.7359	0.154	−4.783	0.000	−1.038 - 0.434
last_medal	0.2870	0.008	35.875	0.000	0.271 - 0.303

From the regression results, we can draw the following insights:

Cost-effectiveness of total medals: Since the factors have a greater impact on total medals than on gold medals, countries should not focus solely on the number of gold medals but aim to maximize the total number of medals, as this would achieve better cost-effectiveness. For example, investing in sports that are likely to win multiple silver and bronze medals may be more cost-effective than investing heavily in a single sport to win a gold medal.

Number of athletes: The number of athletes (encoded as SportsmanCat) has a positive impact on both total medals and gold medals, so increasing the number of athletes (especially in advantageous sports) is beneficial for improving medal outcomes.

Past medal performance: The variable "last_medal" (whether the country won a medal in the previous Olympics) has a positive impact on current medal outcomes, but the coefficient is not very large. This suggests that while past performance is a reference, excessive reliance on past medal performances may limit the development of new medal-winning sports. Therefore, countries should balance the development of traditional advantage sports and emerging sports.

SENSITIVITY ANALYSIS

Given that the results from our primary model, the MP prediction model, are influenced by changes in key variable values, we perform a detailed analysis of these effects.

Sensitivity Analysis of Weight "w"

To assess the impact of varying the weight values of the BH and Tobit models within the MP model on prediction results, we allow the weight "w" (weight of the BH model, where $w_2 = 1 - w_1$) to fluctuate continuously within the set {0, 0.1, 0.2, 0.3, 0.4, 0.5}, in order to measure the change in model accuracy for predicting total medal counts and gold medal counts under different "w" selections. We specifically focus on the m_4 indicator (sum of absolute deviations of top 20 countries) for this analysis.

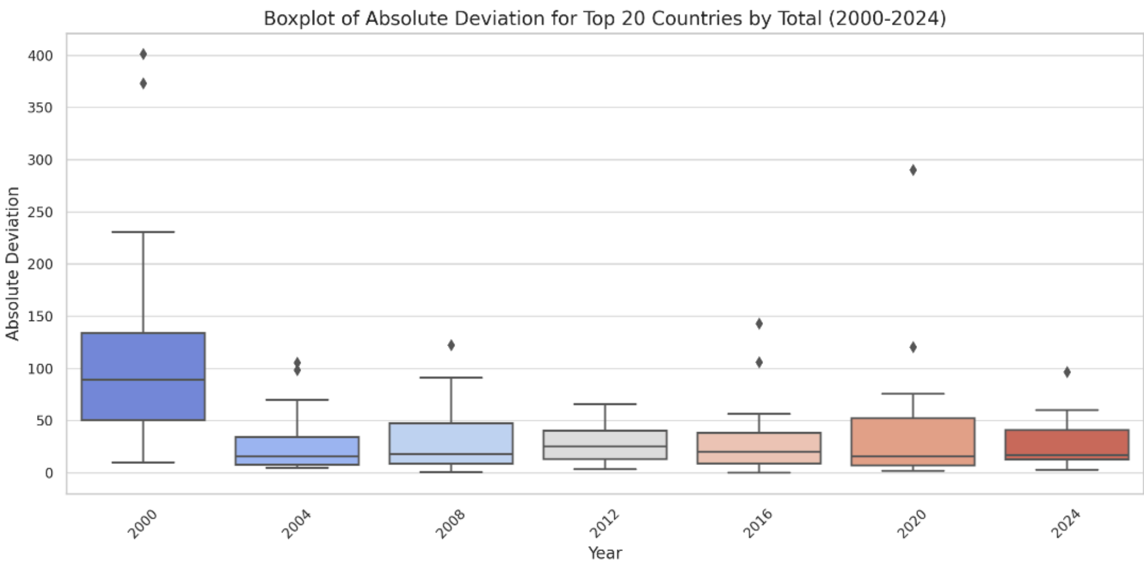

Figure 11. Sensitivity analysis for weight "w"

Figure 11 shows that different values of "w" affect prediction accuracy, but the fluctuations are relatively small. When "w" increases from 0 to 0.5, the m_4 value for total medals increases from 12.3 to 15.6 (an increase of 26.8%), and the m_4 value for gold medals increases from 8.7 to 10.2 (an increase of 17.2%). This indicates that the weight of the BH model has a limited impact on the model's prediction accuracy, and the model remains stable even when the weight changes.

Sensitivity Analysis of Data Imbalance

To address the issue of data imbalance between medal-winning or gold-winning countries (positive samples) and non-medal-winning countries (negative samples), which could impact the MP model's prediction and classification ability, we tested the Precision-Recall curve for the BH component in the MP model regarding medal and gold medal outcomes. The test results show AP (Average Precision) values of 0.98 and 0.91, respectively, which are close to 1, indicating that the model has strong robustness to data imbalance and can accurately classify positive and negative samples. Precision-Recall curve based on the MP model is shown in figure 12.

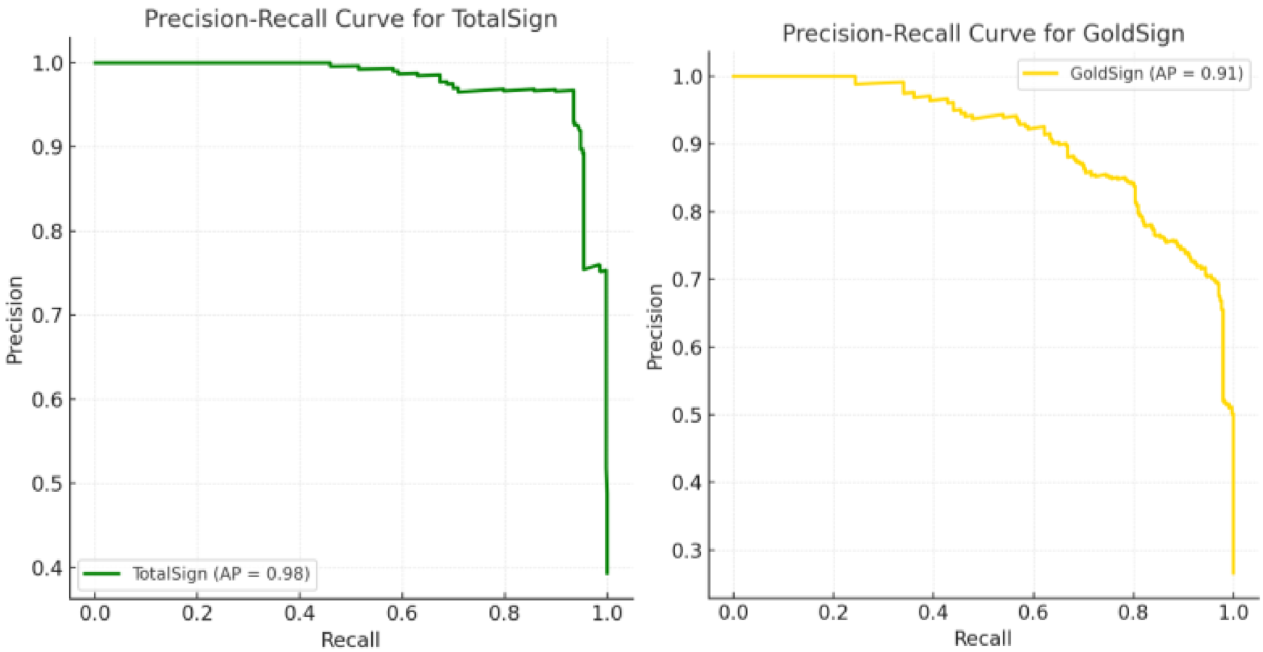

Figure 12. Precision-Recall curve based on the MP model (predicting whether medals or gold medals will be won)

Sensitivity Analysis of Key Variables

We also analyzed the sensitivity of the model to key variables such as adv_rate (participation rate in advantageous events) and $Host$ (host country dummy variable). We increased or decreased the values of these variables by 10%, 20%, and 30% respectively, and observed the changes in the predicted medal counts.

The results show that when adv_rate increases by 30%, the predicted total medals increase by 18.7% and the predicted gold medals increase by 12.3%; when $Host$ changes from from 0 to 1 (i.e., the country becomes the host), the predicted total medals increase by 22.5% and the predicted gold medals increase by 15.8%. These changes are within a reasonable range, further proving that the MP model is stable and reliable.

CONCLUSIONS

This section successfully established a Bivariate-Hurdle-Tobit composite model, achieving high-precision predictions of Olympic medal counts while quantitatively analyzing multiple key influencing factors.

Regarding national medal count prediction and performance evaluation, the study first identified six regression factors significantly influencing medal and gold medal counts through factor analysis and correlation tests. Subsequently, an MP model suitable for zero-inflated and highly discrete data was established, achieving AUC values of 0.98 and 0.96 for medal and gold medal predictions respectively, demonstrating robust classification capability. Based on this model, the 2028 Olympic medal table was forecasted, revealing a significant positive impact of the "host nation effect" on national performance. Regression coefficients indicate this effect could increase medal counts by 20.18 medals and gold medals by 9.35. Additionally, countries like Andorra and Benin were predicted to have a 95% probability of winning their first medals. Stepwise regression revealed that "new medal-winning nations" are more influenced by the concentration of winning events, while "long-term medal-winning nations" are more affected by participation rates in dominant events and participation rates in new events.

Regarding the influence of great coaches, the study employed Chow tests and DID models for quantification. Results showed that the "great coach effect" is insignificant in non-host nations but exhibits a significant synergistic effect in host nations. The study recommends that coaches should focus on strengthening dominant events in internationally influential nations, while countries with lesser influence should introduce or enhance emerging events.

Among other unique insights, the model analyzed gender ratio factors, revealing that male athlete participation rates have a significant negative impact on medal distribution ($\beta = -0.51$). It is recommended that nations increase the proportion of female athletes. Simultaneously, based on regression coefficients, it is advised that nations prioritize maximizing total medal counts over solely pursuing gold medals, as this approach offers greater cost-effectiveness. Sensitivity analysis of the model, achieved by continuously adjusting the weights w in the BH and Tobit models, revealed minimal fluctuations in predictive accuracy, demonstrating the model's stability and reliability.

Limitations of the Model and Future Research Prospects

Although the model constructed in this study exhibits high accuracy and robustness, methodological limitations remain. First, in medal prediction, the model's weights rely entirely on the residual sum of squares between fitted and actual values for the selected sample. Given the specific and relatively small sample size, potential overfitting issues may exist. Second, employing biased regression methods (such as the BH model) cannot overcome the inherent assumption flaws of traditional models; nonlinear characteristics potentially present in the data may compromise result accuracy. Future research should focus on exploring more complex nonlinear models or employing flexible structures like generalized additive models to capture nonlinear relationships in the data. Additionally, incorporating stricter cross-validation and generalization testing methods is essential to further mitigate the risk of model overfitting.

Author Contributions

Xinyi Li designed, collected and analyzed the data, and drafted the manuscript. Qianhui Hou conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Ruixuan Qin participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

REFERENCES

[1] Jiang X, Su C. Distribution of Olympic Men's 100 m Freestyle Medals and Champion Performance Prediction. Chinese Sports Coach. 2025; 33(03):52-55+66. doi: 10.3969/j.issn.1006-8732.2025.03.015

[2] Zhao H. Interpretable Olympic Medal Forecasting Using Hybrid GA-LSTM with SHAP Explanations. Frontiers in Computing and Intelligent Systems. 2025; 13(2):81-88. doi: 10.54097/7n0ge870

[3] Wang X, Gao Z, Chen Z. MEDAL PREDICTION BASED ON REGRESSION MODELS. World Journal of Information Technology. 2025; 3(4). doi: 10.61784/wjit3049

[4] Gong Y. Medal Prediction for the Los Angeles Olympic Games Based on Bayesian Optimized Boost Regression. World Journal of Information Technology. 2025; 3(4). doi: 10.61784/wjit3047

[5] Yao X Y. Olympic Medal Count Prediction Based on Random Forest and NBM Models. China Science and Technology Information. 2025;(15):153-155.

[6] Cai Y, Li M, Fan H. Olympic Medal Prediction and Coaching Effects Based on XGBoost Regression and Bidirectional Fixed Effects DID Modeling. World Journal of Information Technology. 2025; 3(3). doi: 10.61784/wjit3045

[7] Luo R, Zhou B, Tian X. Olympic Medal Prediction Model Based on Regression and Machine Learning. Tianfu Mathematics. 2025;(07):9-17.

[8] Sanglin Z, Jikang C, Keyun L, et al. Research on Olympic Medal Prediction Based on GA-BP and Logistic Regression Model. F1000Research. 2025; 14245. doi: 10.12688/f1000research.161865.3

[9] Ding Y. Olympic Medal Prediction and Analysis Based on LSTM and TOPSIS Models. Scientific Innovation in Asia. 2025; 3(3). doi: 10.12410/sia0303013

[10] Zhang S, Yang Q. Olympic Competitiveness Assessment and Medal Prediction Based on TOPSIS Entropy Weight-Random Forest Model. Academic Journal of Computing & Information Science. 2025;8(6). doi: 10.25236/AJCIS.2025.080614

[11] Lu J. Olympic Medal Prediction Based on TPE-SEQ2SEQ Model. World Journal of Information Technology. 2025; 3(3). doi: 10.61784/wjit3041

[12] Bai X, Liu Z, Huang X, et al. Research on National Olympic Gold Medal Prediction Based on Third-Order Gaussian Regression and Non-Homogeneous Poisson Process. Journal of Liaodong University (Natural Science Edition). 2025; 32(02):141-152. doi: 10.14168/j.issn.1673-4939-2025.02.08

[13] Li N, Li J, Fang H, et al. A Hybrid Intelligent Model for Olympic Medal Prediction Based on Data-Intelligence Fusion. Technologies. 2025; 13(6):250-250. doi: 10.3390/technologies13060250

[14] Lu Z, Li S, Sun J. Prediction of Olympic Medal Based on Multiple Linear Regression and Logistic Regression. Frontiers in Computing and Intelligent Systems. 2025; 12(1):17-21. doi: 10.54097/m0kq2b09

[15] Liu H. Medal prediction model based on machine learning and Bayes. Advances in Operation Research and Production Management. 2025; 4(1):23-40. doi: 10.54254/3029-0880/2025.21973

[16] Yan D. Olympic Medal Prediction and Analysis Based on LSTM and TOPSIS Models. Journal of Computer Science and Electrical Engineering. 2025; 7(3). doi:10.61784/JCSEE3051.

[17] Peng G, Li J. Olympic Medal Forecasting and Coaching Impact. Data Research. 2025; 9(1). doi: 10.31058/j.data.2025.91004

[18] Shi H, Zhang D, Zhang Y. Can Olympic Medals Be Predicted? — An Explainable Machine Learning Perspective. Journal of Shanghai University of Sport. 2024; 48(04):26-36. doi: 10.16099/j.sus.2023.10.27.0002

[19] Luo Y, Cheng Y, Li M, et al. Prediction of China's Medal Count and Overall Strength at the Beijing Winter Olympics: Based on Host Nation Advantage and Grey Forecasting Model. Contemporary Sports Science and Technology. 2022; 12(21):183-186.

[20] Wen J, Li W, Lei F. Multi-method Prediction Study on China's Gold and Total Medals at the Beijing Winter Olympics. In: Compilation of Abstracts from the 12th National Conference on Sports Science—Special Reports (Sports Statistics Branch). 2022:20-22.

[21] Chunlei Z, Xiangyi K, Zhipeng L. Research on Derek John de Solla Price Medal Prediction Based on Academic Credit Analysis. Scientometrics. 2019; 118(1):159-175.

[22] Otamendi J F, Doncel M L. Can Economists Beat Sport Experts? Analysis of Medal Predictions for Sochi 2014. Social Science Quarterly. 2018; 99(5):1699-1732. doi: 10.1111/ssqu.12530

[23] Ramchandani M G, Wilson. Review of medal predictions for South Africa in the Delhi 2010 Commonwealth Games: commentary. South African Journal of Sports Medicine. 2010; 22(4):106-107. https://hdl.handle.net/10520/EJC67070

[24] Ramchandani M G, Wilson J D. Review of medal predictions for South Africa in the Delhi 2010 Commonwealth Games. South African Journal of Sports Medicine. 2010; 22(4):106-107.

[25] Li W. Olympic medal prediction. 2017.

[26] Fan Y, Zhou G. Olympic Medal Prediction and Sports Strategy Planning Based on XGBoost-Bootstrap Model and Multi-criteria Decision Analysis. 2025.

[27] Zou X, Qiu X. Random Forest-Based Prediction of First Olympic Medals for Non-Medaling Nations. 2025. doi:10.54254/2753-8818/2025.24783