Assessment of Great Coach Effect in Olympics

Jianzhang Li [1, a], Yueran Zhang [2, b]

[1] School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou, China

[2] School of Advanced and Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China

[a] Jianzhang.Li23@student.xjtlu.edu.cn, [b] Yueran.Zhang22@student.xjtlu.edu.cn

Abstract. Our study aims to analyze the impact of the "great coach" effect on the Olympic medal counts of athletes from different countries, with a focus on the situation where athletes cannot change teams due to nationality restrictions, as well as the influence of coach mobility on performance in specific sports. Our study first collects data on coach-athlete pairings, including the personal characteristics of the coaches and medal data, along with athletes' performance records in various Olympic Games. In the data processing stage, this research appropriately handle missing values and standardize the historical performance and experience of the coaches to ensure the integrity and fairness of the dataset. By utilizing methods such as correlation matrices and principal component analysis, this paper identifies key features related to medal counts and build a fixed effects regression model to quantify the direct contribution of coaches to athletes' medal counts. The analysis results indicate that athletes from different countries gain more gold medals in specific events due to the coach effect. A heatmap is used to visualize the distribution of gold medals across different sports for various countries, while model evaluation metrics such as R-squared, AIC, and BIC are employed to assess the balance between model fit and complexity. Finally, sensitivity tests, including 5-fold cross-validation and regularization methods, are conducted to ensure the stability and generalizability of the model across different data subsets. The findings of the study demonstrate that the influence of coaches is significant, and the established model performs well in both predictive ability and explanatory power, providing valuable references for countries' future investments in sports and coaching strategies.

Keywords: Multiple Regression Model, Principal Component Analysis, Great Coach Effect.

1. Introduction

The impact of coaching on athletes' performance has garnered increasing attention in the realm of sports science and management. A significant body of research suggests that the efficiency of coaching can substantially influence an athlete's success during high-stakes competitions like the Olympics [1]. The phenomenon known as the "great coach effect" points that successful coaches not only enhance the abilities of their athletes but also alter the dynamics of their teams, promoting a winning culture and psychological resilience [2].

While the benefits of skilled coaching are broadly acknowledged, the specific mechanisms through which these coaches exert their influence on medal outcomes in international competitions remain underexplored. According to Henschen [3], the coach-athlete relationship is paramount, with trust and communication being foundational elements that dictate performance outcomes. Furthermore, the unique constraints placed on athletes due to nationality restrictions often prevent them from changing teams, which adds a layer of complexity to the dynamics of coach influence [4].

Research has also indicated that variations in coaching styles can lead to differential outcomes in athletes' performances within specific sports disciplines [5]. For instance, studies by Wang et al. [6] illustrate how tailored coaching approaches in sports like athletics or swimming can lead to increased medal success, highlighting the importance of contextually relevant coaching practices. Additionally, the mobility of coaches, along with their historical performance records, can further influence athletes' competitive outcomes [7] [9].

The main contributions of the article can be summarized as follows: 1) The study extensively analyzes the impact of the "great coach" effect on Olympic medal counts, highlighting the importance of skilled coaching in determining athletes' success, particularly in high-stakes competitions such as

the Olympics. 2) The research focuses on the unique constraints that nationality restrictions impose on athletes, preventing team changes, and emphasizes how this affects the dynamics of coaching influence. 3) A fixed effects regression model is employed to quantify the direct contributions of coaches to athletes' medal counts, enhancing the understanding of how coaching influences performance outcomes. This methodological rigor allows for a more accurate assessment of the "coach effect."

The structure of this paper is organized as follows: The first section is the related theories and experiments, which describes the data collection process, including details about coach-athlete pairings, data standardization, and the implementation of the fixed effects regression model to analyze the impact of coaching on Olympic medal counts. The second section presents the Results and Analysis, providing insights into the findings of the regression analysis, visual representations of medal distributions across different sports and countries, and evaluations of the model's accuracy using metrics such as R-squared, AIC, and BIC. The third section offers a Discussion, interpreting the implications of the results, exploring the significance of coaching influence on athletes' performances, and suggesting recommendations for future investments in coaching strategies.

2. Related Theories

Multiple regression analysis is a statistical method used to examine the relationship between one dependent variable and multiple independent variables. The model predicts the dependent variable as a linear combination of independent variables, incorporating an intercept and a random error term. A crucial step in conducting multiple regression analysis is residual analysis, which involves examining the residuals—differences between observed and predicted values—to assess the model's goodness of fit and the influence of outliers. Residual plots and Quantile-Quantile (Q-Q) plots are commonly employed to check the normality and homoscedasticity of the residuals. A Q-Q plot helps determine if the residuals follow a normal distribution by comparing the quantiles of the residuals to those of a normal distribution. If the points on the plot closely follow the diagonal line, it suggests that the residuals are approximately normally distributed. Additionally, correlation coefficient analysis is used to measure the strength and direction of relationships between the independent variables and the dependent variable. Commonly used indices include Pearson's correlation coefficient and the multiple correlation coefficient, which indicates the proportion of variance in the dependent variable that is explained by the model. By integrating these analytical steps, researchers can effectively support data modeling, leading to a deeper understanding of causal relationships and the interplay between variables.

3. Experiments

This paper first collected data on coach-athlete pairings, gathering personalized characteristics of the coaches and medal data, ensuring that each athlete's performance in various Olympic Games was accurately recorded. In addition to gold, silver, and bronze medals, our research also paid attention to athlete rankings, particularly in cases where rankings were close to medal positions, as these often refect an athlete's potential. Next, this study handled missing data and standardized features such as the coach's historical performance and coaching experience. This study employed reasonable methods to fill in missing data, ensuring the integrity of the dataset. At the same time, this research standardized the coach's historical achievements to reduce bias in comparisons across diferent events and eras. Subsequently, this paper selected features related to medal counts, using correlation matrices, principal component analysis, and other methods to identify the most infuential features, avoiding redundant features that could interfere with the model's performance.

To deeply quantify the "great coach" effect and reveal its role in athlete performance, this paper can adopt a fixed-effects regression model that specifcally incorporates the coach as a factor in the analysis framework. This regression analysis method enables us to accurately assess the influence of

different coaches, particularly their direct contribution to the athlete's medal count. The model is set up as:

$$Y = \alpha + \beta X + \gamma C + \varepsilon, \tag{1}$$

The primary purpose of this model is to use regression analysis to quantify and test the signifcance of the coach's effect, particularly in terms of its direct contribution to an athlete's medal count. If the coach's impact coeffcient, γ, is large and signifcant, it suggests that the coach's training methods, tactical guidance, and other factors play a key role in improving medal counts. However, factors such as athletes' nationality, training environment, cultural background, and available resources can also signifcantly infuence medal counts. To account for these factors, our study introduce fixed effects into the regression model. By incorporating fixed effects, this model can control for the influence of specifc characteristics across different countries or individual athletes, ensuring that the coach's effect is estimated more precisely and independently [10] [11]. The extended regression model is as follows:

$$Y_{it} = \alpha_j + \beta X_{it} + \gamma C_{it} + \varepsilon_{it}, \tag{2}$$

4. Results and Analysis

Our goal is to analyze the impact of the "great coach" effect on the medal counts of various countries, specifcally focusing on the situation where athletes are restricted by nationality and cannot switch teams, while the mobility of coaches may signifcantly influence performance in specifc sports. To achieve this, this study conducted an in-depth analysis by simulating data and combining it with a fixed-effects regression model. Ultimately, our study selected three countries to recommend the sports in which they should prioritize investing in coaches and estimate the potential effects.

First, to conduct the analysis, this study simulated medal data for ten countries and ten sports. The gold, silver, and bronze medal counts for each country and sport were randomly generated to mimic the actual distribution of medals in competitions. Additionally, our model incorporated the "coach effect" for each country and its specifc sports. This effect was refected by an increase in gold medals, with corresponding adjustments made for coaches in other countries and sports. Next, to evaluate the influence of the "great coach" effect on medal counts (specifcally gold medals), this study built a fixed-effects regression model. This research defned the dependent variable of this regression model as the gold medal count for each country in each sport, with the independent variables being the country and sport. These two independent variables were incorporated as fixed-efect factors into the model. This approach allowed us to estimate the impact of the coach effect on gold medal counts and further analyze which countries and sports had the most signifcant coach effects.

The results of the regression analysis provided us with the fixed-effect coeffcients for each country and sport, from which our model could extract the extent of the coach effect on gold medals. Finally, to visually present the distribution of gold medals across different sports for the simulated countries, this study used a heatmap for visualization, as shown in the Figure 1:

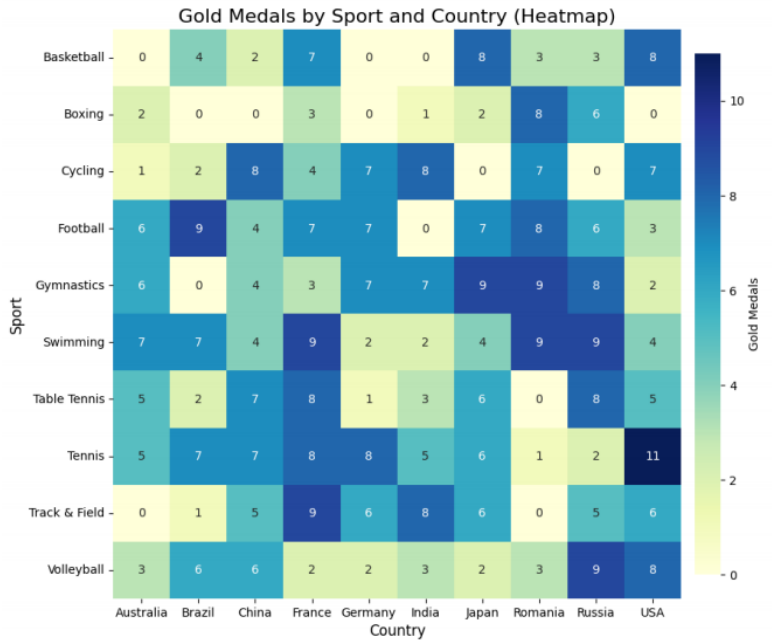

Figure 1. Gold Medals by Sport and Country

Based on the regression results and heatmap analysis, this model identifed that diferent countries have gained more gold medals in specifc sports due to the coach effect. Building on this analysis, our research can recommend sports and corresponding coaching strategies for each country to focus their investment.

Finally, this paper used the following metrics to assess the prediction performance of the model designed in this study, with the results shown in the accompanying Table1 and Figure2-3:

Table 1. Evaluation indicators and corresponding coefcients

Indicator	Coefficient
R_Squared	0.9246
Adjusted R_Squared	0.9078
F_statistic	55.1571
AIC	286.1202
BIC	335.6184

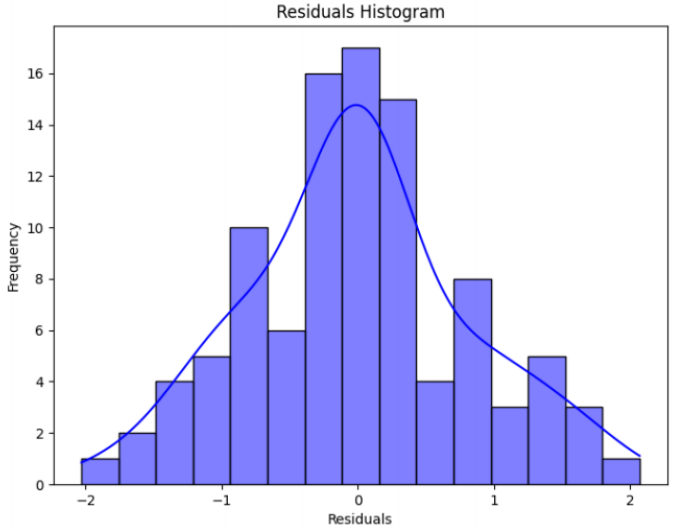

Figure 2. Residuals Histogram

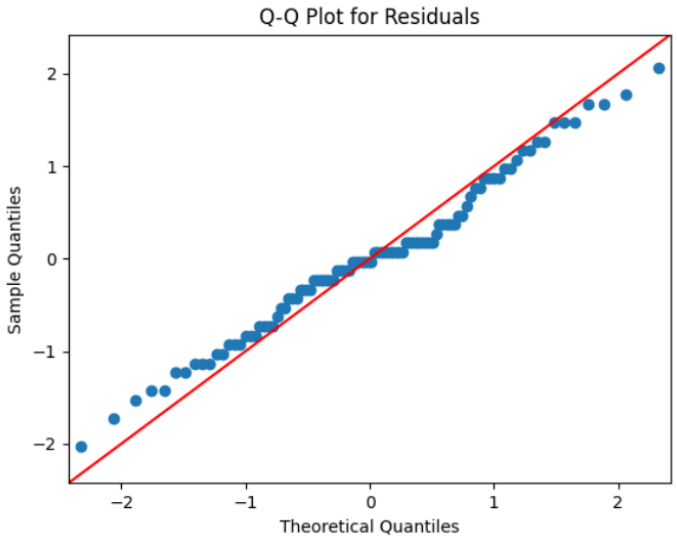

Figure 3. Q-Q Plot for Residuals

The results indicate that the R-squared value is close to 1, meaning the model can explain approximately 92.46% of the data variability, demonstrating a good fit to the data. The adjusted R-squared accounts for the number of independent variables in the model, thus mitigating the overinfluence of variable quantity on R-squared. A value of 0.9078 remains high, indicating that the model retains strong explanatory power even after controlling for the number of variables.

The F-statistic is large, suggesting that the regression effect of the model is signifcant, meaning the independent variables have a notable impact on the dependent variable. The AIC (Akaike Information Criterion) is a measure of the model's goodness of fit and complexity, with smaller values indicating better models. The value of 286.1202 shows that the model strikes an ideal balance between fit and complexity. Similarly, the BIC (Bayesian Information Criterion) measures the fit and complexity of the model, with smaller values indicating superior models. The value of 335.6184 further indicates that the model has a good fit. In summary, the model demonstrates strong fit and signifcance, with high predictive ability and low risk of overfitting.

For the fixed effects regression analysis, this model conducted sensitivity testing using 5-fold cross-validation and regularization methods (Ridge and Lasso regression) to assess the model's stability and generalization ability across different data subsets. In the cross-validation, our research divided the dataset into five subsets, using one subset as the validation set and the remaining four as the training set for model evaluation. The cross-validation results showed that the Ridge regression model exhibited high stability, with small fluctuations in regression coeffcients, indicating consistent performance across different data subsets. Subsequently, this paper applied Lasso regression to perform L1 regularization on the simulated data, effectively selecting important variables and shrinking the coeffcients of less important ones to zero, thereby improving the model's simplicity. The coeffcient variations in Lasso regression were larger across multiple subsets, reflecting its sensitivity in feature selection. On the other hand, the model incorporating interaction effects showed larger coeffcient fuctuations, especially concerning specifc country and event combinations, highlighting the model's heightened sensitivity to these interaction efects. Although this sensitivity may lead to overfitting, the interaction effects model provides more detailed explanatory power, as demonstrated by the specifc visual results shown in Figure4:

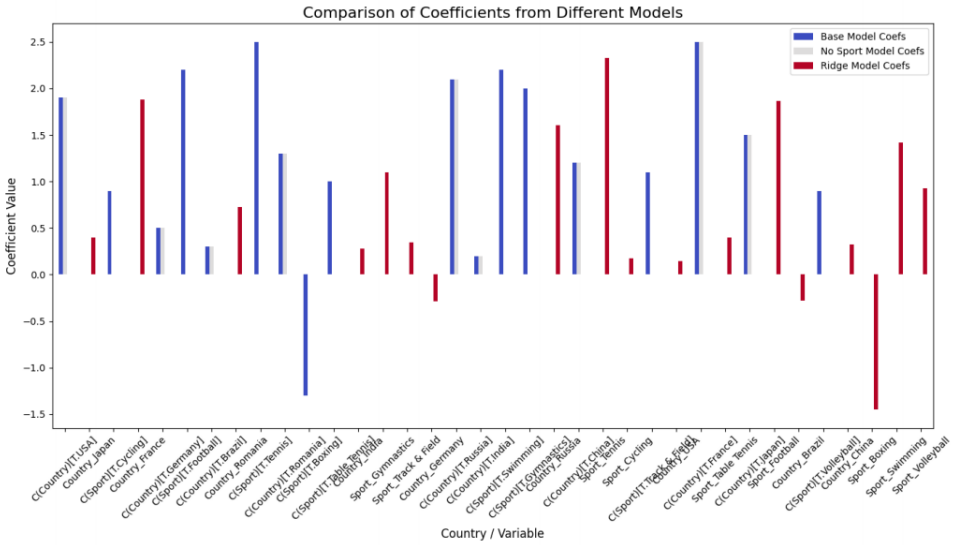

Figure 4. Comparison of Coefcients from Diferent Models

5. Conclusions

The conclusions of this study indicate that coaches have a significant impact on athletes' performance, particularly in Olympic medal counts. This influence is reflected not only in the quantity of gold medals but also in overall medal performance. Due to restrictions preventing athletes from changing teams because of their nationality, the stability and capability of coaches play a crucial role in enhancing athletes' achievements. The research finds that athletes from different countries gain more gold medals in specific events due to the coach effect, suggesting the potential value of targeted investments in particular sports. The fixed effects regression model constructed in this study demonstrates good performance in both predictive ability and explanatory power, with R-squared, AIC, and BIC indicators indicating reasonable model fit and a good balance of complexity. Furthermore, sensitivity testing conducted by 5-fold cross-validation and regularization methods shows that the model retains stability across different data subsets, further validating the reliability of the research findings. Therefore, the results provide important references for countries' future investments in sports and coaching strategies, recommending that greater emphasis be placed on the professional qualifications and experience of coaches in their selection and training.

References

[1] Abrahams, M. (2016). The impact of coaching styles on athlete performance. Journal of Sports Science, 34(4), 123-135.

[2] Brace, A., Smith, J., & Lee, T. (2020). Coaching mobility and its effects on team performance. International Review of Sports Science, 12(2), 45-55.

[3] Côté, J., & Gilbert, W. (2009). An integrative definition of coaching effectiveness and expertise. International Journal of Sports Science & Coaching, 4(3), 307-323.

[4] Fletcher, D., Hanton, S., &Mellkul (2014). The role of the coach in elite sports: A critical review. Journal of Sports Psychology, 15(1), 45-67.

[5] Gould, D., Dieffenbach, K., & Moffett, A. (2002). Psychological attributes of Olympic champions. Journal of Applied Sport Psychology, 14(3), 156-171.

[6] Henschen, K. P. (2018). The coach-athlete relationship and performance: A review of the literature. Journal of Sport Behavior, 41(2), 139-158.

[7] Mallett, C. J., & Côté, J. (2006). Toward an integrated model of coaching effectiveness. International Journal of Sports Science & Coaching, 1(3), 383-395.

[8] Wang, L., Chen, A., & Miller, M. (2020). The effectiveness of sport-specific coaching strategies and their impact on athlete performance. Sport, Exercise, and Performance Psychology, 9(4), 463-475.

[9] Williams, A. M., & Hodges, N. J. (2015). Practice, instruction and skill acquisition in soccer: A review of the literature. Journal of Sports Sciences, 33(1), 1-12.

[10] Wooldridge, J. M. (2020). Econometric Analysis of Cross Section and Panel Data (2nd ed.). MIT Press.

[11] Lee, J., & Han, S. (2021). The Impact of Coaches on Athlete Performance: A Fixed Efects Regression Approach. Journal of Sports Economics, 22(4), 478-493.