Insights of Economics System
WISDOM ACADEMIC ISSN: 3080-731X | E-ISSN: 3080-7328 Volume 2 , Issue 1

Article

An Interpretable, Year-Grouped Machine-Learning Framework for Forecasting Olympic Medal Counts with Uncertainty Quantification: Evidence from 1960–2024 and Projections for LA 2028

Jiaxin Zhang[1], Luping Tang[1,*]
[1]University of South China, Hengyang 421001, China.
[*]**Corresponding author:** Luping Tang, tanglp99@163.com.

CITATION

Zhang JX & Tang LP. An Interpretable, Year-Grouped Machine-Learning Framework for Forecasting Olympic Medal Counts with Uncertainty Quantification: Evidence from 1960–2024 and Projections for LA 2028. Insights of Economics System. 2026; Vol 2 (No. 1): 313.

https://doi.org/10.63808/ies.v2i1.313

ARTICLE INFO

Received: 16 January 2026
Accepted: 19 January 2026
Available online: 29 January 2026

Abstract: Olympic medal totals are highly right-skewed and concentrated among a small group of top-performing nations, which complicates stable cross-country forecasting and interpretation. Using an unbalanced country-by-edition panel for the 1960 to 2024 Summer Games, we build a reproducible pipeline to predict national total medals and to examine plausible drivers. The target is log-transformed during training and then converted back to the medal scale with a non-negativity constraint. Predictors include host-country status, pre-Games macroeconomic conditions (gross domestic product, population, and gross domestic product per capita), performance inertia from the previous Olympic edition, and participation-related proxies such as athlete counts and coverage of sports and events when available. Random Forest and histogram-based gradient boosting models are evaluated using cross-validation that groups observations by Olympic year to better reflect next-edition forecasting and to limit within-year information leakage. For the Los Angeles 2028 Games, we construct features from the 2024 baseline and extrapolate macroeconomic inputs to the pre-Games year using country-specific compound growth rates. Uncertainty is quantified with bootstrap prediction intervals. Results indicate that the United States and China remain the leading medal producers, while interval overlap suggests that rankings among mid-tier nations are sensitive to uncertainty. Feature attribution highlights participation intensity and opportunity breadth as consistent positive contributors, whereas host status shows a smaller marginal role after controlling for observable factors. An event-study difference-in-differences specification yields imprecise host-effect estimates, underscoring the need for cautious causal interpretation.

Keywords: Olympic medal forecasting; interpretable machine learning; SHAP; year-grouped cross-validation; bootstrap prediction intervals; host-country effect; event-study difference-in-differences; panel data; Random Forest; Histogram-based Gradient Boosting

1. Introduction

Olympic medal outcomes are not only a concentrated display of competitive performance but are also widely used as an observable indicator of a country's capacity for sport governance, resource allocation, and the efficiency of talent development systems. Regarding the questions of which factors drive medal production and how to make reproducible forecasts for the next edition, prior studies have largely adopted a cross-national comparative perspective and emphasized the fundamental constraints imposed by macroeconomic strength and population size. In general, greater economic resources and a larger potential population base imply better training provision, broader talent selection, and higher long-run competitive output (Bernard and Busse, 2004; Grimes et al., 1974; Johnson and Ali, 2004; Lui and Suen, 2008). Within this line of inquiry, gross domestic product, population, and related derived indicators have repeatedly been shown to exhibit systematic associations with medal performance, and they have gradually become the macro-level foundation in a wide range of forecasting models (Lozano et al., 2002).

Beyond resource endowments, the host effect is among the most policy-relevant phenomena in Olympic medal research. Host nations may hold potential advantages through quota allocation, environmental familiarity, spectator support, and organizational mobilization, thereby generating additional medal gains in the hosting edition and adjacent editions (Balmer et al., 2003). However, host effects are deeply intertwined with long-run national strength, cyclical investment patterns, and adjustments in sport-specific structures. If examined only through cross-sectional comparisons or simple correlational frameworks, it is easy to conflate hosting-related gains with underlying time trends driven by pre- and post-host investment surges, which can lead to biased policy interpretation (Breiman, 2001). Therefore, alongside forecasting, introducing a dynamic evaluation framework that is closer to causal identification to test and characterize trajectories before and after hosting is essential for improving both interpretability and practical usefulness.

Methodologically, for structured panel data at the national and Olympic-edition level, tree-based models are well suited to capturing nonlinearities and interaction effects and have become a common choice in multivariate prediction tasks. Random Forest improves generalization robustness by averaging across an ensemble of regression trees, thereby reducing variance (Breiman, 2001). Gradient boosting models build strong predictive capacity by iteratively fitting residuals and often achieve superior fit and extrapolative performance in complex feature spaces (Friedman, 2001). At the same time, increasing model complexity raises concerns about black-box behavior. If a model cannot answer which factors drive predictions and which variables contribute to the forecast for a specific country, its outputs are difficult to translate into actionable policy implications. Shapley-value-based interpretability offers a unified route to address this issue: it can rank influential features at the global level and decompose feature contributions for individual observations, integrating prediction and explanation within a single analytical (Lundberg and Lee, 2017; Pedregosa et al., 2011; Shapley, 1953).

Building on this research trajectory, this study constructs a panel dataset for the Summer Olympics from 1960 to 2024 at the level of National Olympic Committee and Olympic edition, with total medal counts as the prediction target. We adopt a feature-alignment strategy consistent with real-world preparation cycles. Macroeconomic and demographic indicators are aligned to the year preceding each edition to reflect the lagged influence of resource conditions on preparation and participation capacity. Historical performance and participation-scale variables are used to represent competitive inertia and the intensity of delegation investment. Given the pronounced right-skewness of medal distributions and the presence of extreme values, the target variable is log-stabilized during model training and then mapped back to the original medal scale at output, which mitigates the dominance of extreme observations from top nations in the loss function and improves overall predictive stability across countries. Model evaluation follows a cross-validation design grouped by Olympic year to better approximate the applied setting of forecasting the next edition from prior editions while reducing the risk of within-edition information leakage.

On the output side, we not only provide point forecasts for the 2028 Los Angeles Games but also quantify uncertainty by constructing a predictive distribution via bootstrap resampling and reporting percentile-based intervals, thereby enhancing usability for risk communication, scenario analysis, and uncertainty reporting. We further apply interpretable machine learning to decompose feature contributions at both the global and country-specific levels. Finally, we employ an event-study difference-in-differences framework with two-way fixed effects to evaluate mechanisms by tracing the dynamic path of host effects across multiple editions before and after hosting. We remain cautious about potential estimation bias in multi-period difference-in-differences settings under heterogeneous treatment timing, and, when necessary, we suggest complementary checks using more robust identification approaches for staggered treatment effects. Overall, the study develops an integrated paradigm that combines reproducible forecasting, uncertainty quantification, interpretable attribution, and quasi-causal testing, offering a framework for Olympic medal prediction that is usable, explainable, and less prone to misinterpreting correlations as causal effects.

2. Data

2.1. Data sources

This study constructs a Summer Olympics panel dataset at the country or region by edition level. The observational unit is defined by the National Olympic Committee code (National Olympic Committee, NOC) together with the Olympic edition year (year). The data are drawn from three main sources. First, Olympic competition and medal information is based on the modern Olympic historical dataset released through Tidy Tuesday, an open data initiative designed for data science practice. The dataset contains fields on athletes, events, and medals, which enables aggregation by NOC and year to obtain gold medals (gold), silver medals (silver), bronze medals (bronze), and total medals (total). When detailed fields are available with consistent definitions, the dataset further supports the construction of proxy variables describing delegation size and event participation breadth. Second, macroeconomic and demographic data include gross domestic product (Gross Domestic Product, GDP, measured in current United States dollars, USD) and population (population)(Efron, 1979). These data are obtained from publicly available, standardized comma-separated values files (Comma-Separated Values, CSV) that have been curated into a unified structure. The compilation follows the statistical definitions and reporting standards of the World Bank (World Bank, WB), which facilitates alignment with Olympic editions and subsequent merging. Third, host-country information is used to construct the host indicator and to define the event-study window. Hosting status is compiled from the official Olympic website (Olympics.com), including confirmed host city and edition information, together with the Los Angeles 2028 (LA 2028) pages, to ensure authoritative identification and consistent mapping between hosting editions and countries.

2.2. Sample coverage and variable definitions

In terms of sample coverage, the analysis uses an unbalanced panel at the national level, where imbalance arises because not all countries participate in every edition and not all countries win medals in each edition. Countries or regions are identified by the National Olympic Committee code (National Olympic Committee, NOC; variable name noc), and Olympic editions are indexed by year (year). For model construction, priority is given to core variables that are consistently defined across editions and can be obtained reliably over time. These include medal outcomes (gold, silver, bronze, total), host status (is_host), historical performance inertia variables (prev_*, indicating medal counts in the previous Olympic edition), and macroeconomic constraints (gdp_usd, population, gdp_per_capita_usd). Macroeconomic variables follow a pre-edition alignment rule, implemented by defining the macro alignment year as the year preceding the Olympic edition, namely macro_year equals year minus one. This alignment is intended to make GDP and population better reflect pre-Games resource endowments and preparation capacity, thereby reducing the risk that post-Games accounting or contemporaneous measurement mechanically reflects the competition outcome (Sun and Abraham, 2021). In addition, delegation size and participation breadth variables, including athlete count (athlete_count), sport coverage (sport_count), and event coverage (event_count), are treated as auxiliary explanatory enhancements. They are constructed and included only when the underlying sources provide detailed records with consistent definitions, in order to minimize systematic bias arising from differences in statistical standards across data sources.

2.3. Data cleaning, aggregation, and feature engineering

The data processing follows a structured workflow comprising field harmonization, restriction to the Summer Games, medal aggregation, macro-level alignment, construction of lagged variables, and final panel export. We first standardize raw fields to ensure the presence of key columns including noc, country, year, and medal, and, when a season field is available, retain only records corresponding to the Summer Olympics. Next, non-missing medal records are counted and aggregated by noc, country, year, and medal, then reshaped into gold, silver, and bronze, with total medals defined as the sum of gold, silver, and bronze. This yields a country-by-edition medal panel.

When consistent athlete- and event-level details are available, we further derive participation-related proxies. Specifically, we compute delegation size as the number of unique athletes by noc and year (athlete_count). We also quantify the breadth of participation opportunities by counting distinct sports (sport_count) and distinct events (event_count) by noc and year, which serve as approximations of the opportunity set and the coverage of competitive disciplines.

Macroeconomic variables are then merged using the country three-letter code (ISO 3166-1 alpha-3, ISO3) and the macro alignment year (macro_year), and gross domestic product per capita is calculated as gross domestic product divided by population (gdp_per_capita_usd equals gdp_usd divided by population). To improve cross-source matching, ISO3 codes are prioritized, and a targeted mapping is applied to reconcile the small subset of entities for which the National Olympic Committee identifier does not coincide with ISO3 conventions. Finally, lagged medal variables are created by shifting medal counts back by one Olympic cycle, defined as four years, resulting in prev_gold, prev_silver, prev_bronze, and prev_total. Host status (is_host) is assigned based on hosting editions, for example, the United States is coded as the host for Los Angeles 2028, producing a structured feature set suitable for prediction, interpretation, and subsequent causal diagnostics.

Cross-source integration and historical changes in geopolitical entities can generate missing values in macro variables and participation proxies. For instance, some historical National Olympic Committee entities cannot be matched reliably to World Bank ISO3 entities, leading to missing records for gross domestic product and population. In addition, if certain editions provide only country-level medal totals without detailed athlete or event information, athlete_count, sport_count, and event_count may be unavailable. To maintain model trainability and robustness, continuous features are imputed using median imputation during the machine learning stage, thereby reducing sensitivity of estimation and prediction to missingness. For the event-study difference-in-differences regression used to evaluate host effects, control variables entering the regression are cleaned explicitly through imputation or deletion, and the clustering group vector is constructed to correspond exactly to the final regression sample. This alignment ensures the validity of cluster-robust standard error computation.

To support replication and extensibility, all data processing and modeling steps are implemented as a scripted pipeline. Public raw data are retrieved by script and processed locally through unified cleaning and feature construction. The final country-by-edition panel is exported in comma-separated values format and as an Excel workbook. In addition, model comparisons, backtesting predictions, the Los Angeles 2028 forecasts together with uncertainty intervals, feature attribution visualizations, and event-study host-effect figures are saved into a timestamped results directory. This design provides traceability across experimental versions and enables repeated execution under a consistent data definition to verify key findings.

3. Methods

3.1. Overall framework

This study models Olympic medal production at the country–edition level using an unbalanced panel indexed by National Olympic Committee (National Olympic Committee, NOC) i and Olympic year t. The prediction target is the total number of medals y_{it} (code: TARGET="total"). The workflow follows a reproducible "feature construction → preprocessing pipeline → model training and validation → 2028 scenario prediction → uncertainty quantification → interpretability and causal assessment" structure. Specifically, predictive models are trained on historical editions with observed medals, validated by year-grouped cross-validation to mimic

Insights of Economics System
WISDOM ACADEMIC ISSN: 3080-731X | E-ISSN: 3080-7328 Volume 2 , Issue 1

"predicting a future edition from past editions," then applied to the constructed LA 2028 feature set; uncertainty is quantified via bootstrap resampling; finally, model explanation (SHAP or permutation importance) and a causal event-study difference-in-differences design are used to interpret drivers and isolate dynamic host effects (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021).

3.2. Target transformation and inverse mapping

Olympic medals are highly right-skewed with large cross-country dispersion; directly modeling y_itcan lead to heteroscedastic residuals and overemphasis on extreme values. Consistent with the code setting USE_LOG1P=True, we transform the target as

$$\tilde{y}_{it}=\log(1+y_{it}) \tag{1}$$

Models are trained to predict $\hat{\tilde{y}}_{it}$ on the log scale, and predictions are mapped back to the original medal scale using

$$\hat{y}_{it}=\exp(\hat{\tilde{y}}_{it})-1, \hat{y}_{it}\leftarrow\max(\hat{y}_{it},0) \tag{2}$$

which corresponds to np.expm1() and np.maximum (pred_raw, 0) in the code, ensuring non-negative medal outputs.

3.3. Feature set and alignment rules

For each country–edition observation, we construct a feature vector x_itfrom variables that are stable and broadly available across countries and time (code: feature_cols). The baseline feature set includes: (i) host indicator is_host, (ii) macro constraints—GDP in current USD (gdp_usd), population (population), and GDP per capita (gdp_per_capita_usd), (iii) performance inertia—previous-edition medals (prev_total, prev_gold, prev_silver, prev_bronze), (iv) participation scale and coverage proxies—athlete count (athlete_count), number of sports (sport_count), and number of events (event_count) when available, and (v) a long-run trend term year. Macro variables are aligned to the pre-competition resource base using a one-year lead convention (code: macro_year = year - 1), i.e.,

$$\text{macro\_year}=t-1, \text{gpc}_{i,t-1}=\frac{\text{gdp}_{i,t-1}}{\text{pop}_{i,t-1}} \tag{3}$$

Performance inertia features use the previous Olympic edition (four years earlier), consistent with the code logic that shifts by edition:

$$\text{prev\_total}_{it}=y_{i,t-4},\ \text{prev\_gold}_{it}=\text{gold}_{i,t-4},\ \text{etc.} \tag{4}$$

The host indicator is defined as

$$\text{is\_host}_{it}=\begin{cases}1, & \text{if country } i \text{ hosts edition } t,\\ 0, & \text{otherwise.}\end{cases} \tag{5}$$

3.4. Preprocessing and pipeline implementation

Cross-source merging inevitably introduces missing values in macro variables and participation proxies. To ensure stable training for all models under missingness, we apply median imputation to all numeric features (code: SimpleImputer (strategy="median")) within a unified preprocessing–model pipeline (code: ColumnTransformer + Pipeline). Let I(·)denote the median-imputation operator; the effective model input is z_it=I(x_it).

This design enforces consistent preprocessing across models and across cross-validation folds, avoids leakage from test folds into training statistics, and makes the full modeling procedure end-to-end reproducible.

3.5. Predictive models: Random Forest and Histogram Gradient Boosting

To balance predictive accuracy and robustness on tabular, mixed-scale national indicators, we consider two tree-based regressors implemented in scikit-learn: Random Forest regression (Random Forest Regressor, RF) and Histogram-based Gradient Boosting regression (HistGradientBoostingRegressor, HGB). RF reduces variance via bagging an ensemble of regression trees:

$$\hat{y}_{it}=\frac{1}{B}\sum_{b=1}^{B} T_b\,(z_{it}) \tag{6}$$

where T_b (·)is the b-th tree and Bcorresponds to n_estimators in the code. HGB fits an additive model through iterative boosting:

$$\hat{y}_{it}=\sum_{m=0}^{M} v\, f_m(z_{it}) \tag{7}$$

where $f_m(\cdot)$ is the weak learner at iteration m , v is the learning rate (learning_rate), and Mis the number of boosting iterations (max_iter). Both models operate on the same imputed feature matrix z_{it}, enabling a controlled comparison under identical data processing.

3.6. LA 2028 scenario construction and macro extrapolation

To generate LA 2028 predictions, we construct a 2028 feature table $x_{i,2028}$ using 2024 as the most recent observed baseline (code: base_year=2024, make_2028_rows). The host indicator is set by

$$\text{is\_host}_{i,2028}=\mathrm{I}(\text{NOC}_i=\text{USA}) \tag{8}$$

and inertia terms directly inherit 2024 observed medals:

$$\text{prev\_total}_{i,2028}=y_{i,2024},\text{prev\_gold}_{i,2028}=\text{gold}_{i,2024},\text{etc} \tag{9}$$

Participation proxies (athlete_count, sport_count, event_count) are carried forward from 2024 when available (an inertia approximation consistent with the code). Since macro variables are aligned to $t-1$, 2028 requires 2027 macro values; the code extrapolates GDP and population using the compound annual growth rate (CAGR) derived from the most recent two macro-observations for each country (code: extrapolate ()), i.e.,

$$\text{CAGR}=\left(\frac{v_2}{v_1}\right)^{\frac{1}{y_2-y_1}}-1,\hat{v}_{2027}=v_2(1+\text{CAGR})^{(2027-y_2)} \tag{10}$$

and then computes predicted GDP per capita $\widehat{\text{gpc}}_{i,2027}=\widehat{\text{gdp}}_{i,2027}/\widehat{\text{pop}}_{i,2027}$, matching gpc_pred. appends (gdp_2027/pop_2027). The trained pipeline is finally applied to the 2028 feature matrix, and outputs are mapped back using Eq. (2).

3.7. Uncertainty quantification via bootstrap prediction intervals

To quantify the sensitivity of 2028 forecasts to sampling variation in historical data, we construct nonparametric prediction intervals via bootstrap resampling (code: bootstrap intervals). Specifically, we draw B bootstrap samples from the training set with replacement (code: rng. integers (0, n,size=n)), refit the selected model pipeline each time, and obtain a bootstrap predictive distribution $\left\{\hat{y}_{i,2028}^{(b)}\right\}_{b=1}^{B}$. The q_1 and q_2 quantiles provide a central prediction interval:

$$\hat{y}_{i,2028}^{\text{lo}}=Q_{q_1}(\{\hat{y}_{i,2028}^{(b)}\}_{b=1}^{B}),\hat{y}_{i,2028}^{\text{hi}}=Q_{q_2}(\{\hat{y}_{i,2028}^{(b)}\}_{b=1}^{B}) \tag{11}$$

where the default in the code is $(q_1,q_2)=(0.05,0.95)$ and results are exported as pred_lo_90 and pred_hi_90. This interval reflects the empirical variability induced by re-sampling the historical panel and is reported alongside point forecasts.

3.8. Interpretability and causal assessment of the host effect

To interpret which covariates, drive predictions, we follow the code's two-path strategy in explain_features: when SHAP (SHapley Additive exPlanations, SHAP) is available, we compute additive attributions for the trained tree model on the imputed feature matrix (the code explicitly constructs a DataFrame with columns=feature_cols to ensure feature alignment and disables additivity checks via check_additivity=False for numerical stability). The SHAP decomposition is

$$\hat{\tilde{y}}(z)=\phi_0+\sum_{j=1}^{p}\phi_j(z) \tag{12}$$

where $\phi_j(z)$ is the contribution of feature j to the log-scale prediction for sample z.

The global summary plot is produced by shap. summary_plot, and the code additionally generates SHAP dependence plots for athlete_count, event_count, and prev_total to reveal nonlinearity and interaction patterns (files shap_dependence_*.png). Moreover, the code performs sample-level decomposition through SHAP waterfall plots for two representative NOCs (USA and CHN) using baseline-year (2024) feature rows, thereby showing how each covariate pushes the prediction above or below the expected value for those countries. If SHAP is not installed, we fallback to permutation importance (Permutation Importance, PI) computed by permutation importance, quantifying importance as performance deterioration after randomly permuting feature j:

$$\mathrm{Imp}(j)=\mathrm{E}[L(f,Z,\tilde{y})]-\mathrm{E}[L(f,\pi_j(Z),\tilde{y})] \tag{13}$$

where $\pi_j(\cdot)$ permutes the j-th feature and L denotes the loss/score used internally by scikit-learn. Finally, to move from predictive associations toward event-driven inference on "host advantage," the code implements an event-study difference-in-differences regression (event-study Difference-in-Differences, event-study DiD) in eventstudy_host_did. It defines event time $k=\mathrm{round}((t-T_i)/4)$ for a host country whose host year is T_i (code: k=((year-host_year)/4).round().astype(int)), constructs dummies for $k\in\{-3,-2,0,1,2,3\}$ with $k=-1$ omitted as the baseline, and estimates a two-way fixed-effects model:

$$\tilde{y}_{it}=\alpha+\sum_{k\in\{-3,-2,0,1,2,3\}}\beta_k D_{it}^{k}+\gamma^{\top} w_{it}+\mu_i+\lambda_t+\varepsilon_{it} \tag{14}$$

where μ_i and λ_t are NOC and year fixed effects (code: C(noc)+C(year)), w_{it}includes available controls among prev_total, gdp_per_capita_usd, and population, and standard errors are clustered by NOC (code: cov_type="cluster" with aligned groups). For interpretation, coefficients are transformed into percentage effects on the medal scale as

$$\text{Effect}_k=\left(\exp(\beta_k)-1\right)\times 100\% \tag{15}$$

which corresponds to pct=(np.exp(beta)-1) *100 and is visualized in eventstudy_host_effect.png. Pre-event coefficients (e.g., $k=-3,-2$) provide an empirical check for pre-trends, while post-event coefficients ($k\geq 0$) characterize the dynamic host effect path.

4. Results and Discussion

Figure 1 presents the out-of-fold fit obtained under year-grouped cross-validation. Overall, the scatter points concentrate around the forty-five-degree line, indicating that the model is able to recover the dominant structural relationship governing total medal counts even under the constraint of cross-edition extrapolation. A clear heterogeneity is also observed: the fit is relatively tight in the low-medal range, where most nations are located, whereas dispersion increases markedly among high-medal countries. This pattern suggests that medal production for elite nations is influenced by additional determinants that are not fully captured by macroeconomic covariates and lagged performance alone, such as discipline-specific competitive structure, roster composition, host-related investment, and cyclical fluctuations, which jointly contribute to higher predictive variance at the top end of the distribution. **Figure 2** shows the residual distribution. Residuals are centered around zero and are concentrated in the main mass, implying that systematic bias is broadly controlled. Nevertheless, the distribution exhibits a long tail. On the one hand, a small number of observations display large positive residuals, meaning realized outcomes substantially exceed model predictions; on the other hand, there is also a negative tail associated with overprediction. This asymmetry and tail behavior is consistent with the greater volatility observed in editions involving top-performing nations and host countries. Accordingly, the results for 2028 should be interpreted as conditional forecasts given the available features and historical regularities, and uncertainty should be communicated using interval estimates rather than point predictions alone.

Figure 1

OOF fitting

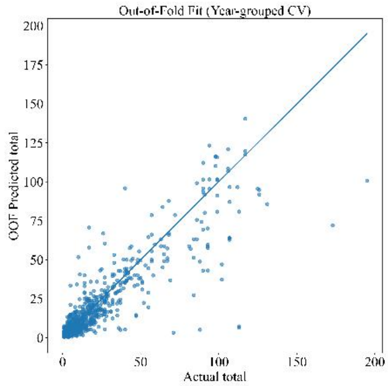


Figure 2

OOF residual plot

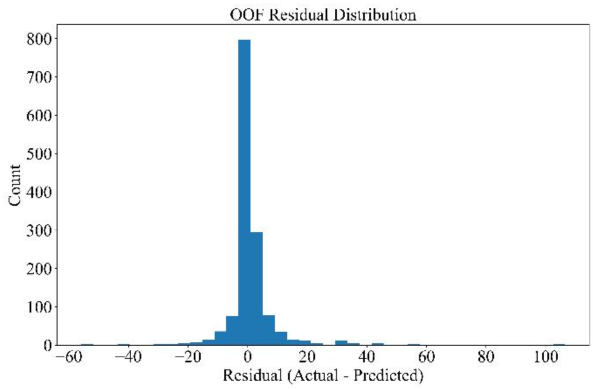
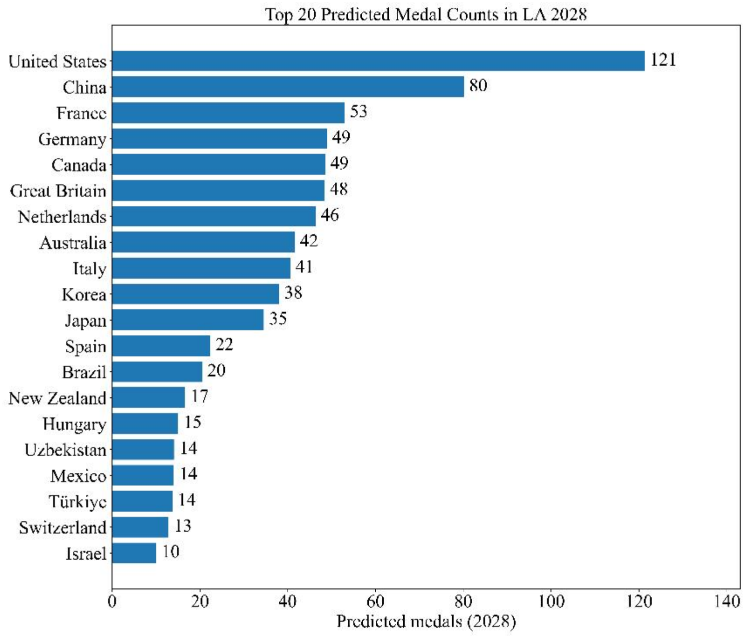


Figure 3 reports the top twenty-point forecasts of total medals for the Los Angeles 2028 Summer Games, measured in medal counts. The predictions place the United States first with an estimated 121 medals, followed by China with approximately 80 medals. A compact second tier is formed by France with 53 medals, Germany and Canada with 49 medals each, the United Kingdom with 48 medals, and the Netherlands with 46 medals, reflecting a distribution characterized by concentration among the leading nations and dense clustering in the upper-middle range. A third tier includes Australia with 42 medals, Italy with 41 medals, the Republic of Korea with 38 medals, and Japan with 35 medals. Beyond this group, Spain with 22 medals, Brazil with 20 medals, New Zealand with 17 medals, and Hungary with 15 medals fall into a moderate-output range. Uzbekistan, Mexico, and

Türkiye are each projected at roughly 14 medals, followed by Switzerland with 13 medals and Israel with 10 medals, forming the lower tail among the top twenty. Overall, this ranking pattern is consistent with the combined mechanism implied by the model specification, namely performance inertia from the previous edition, participation scale and opportunity breadth, and macro-level resource constraints. Nations at the top benefit from pronounced advantages in lagged medal outcomes and participation proxies, which supports the persistence of high predicted totals. In contrast, differences among mid-ranked countries are relatively small, making their ordering more sensitive to marginal shifts in covariates and, consequently, more exposed to ranking uncertainty under plausible feature variation.

Figure 3

Top 20-point forecast results for the total number of medals in LA 2028



Relying solely on point forecasts may obscure heterogeneity in the model's confidence across countries. Therefore, **Figure 4** reports the ninety percent prediction intervals for the top twenty nations, constructed from bootstrap quantiles. The results show that leading countries such as the United States and China, despite having higher point predictions, still exhibit clearly nontrivial interval widths, suggesting that medal production among elite nations is more exposed to volatile factors that are not fully captured by the observable covariates. Notably, several countries in the mid-medal range also display wide intervals. This pattern is consistent with greater edition-to-edition fluctuations in their historical trajectories and with measurement uncertainty arising from incomplete covariates. For example, missing participation proxies in some editions or gaps in macroeconomic series that require imputation or extrapolation can increase dispersion in the predictive distribution. This finding is important for interpretation and reporting. The interval width can be treated as an operational measure of forecast risk, implying that rankings among mid-tier nations should not be over-interpreted. Instead of focusing on the exact ordering of point estimates, greater weight should be placed on whether prediction intervals overlap, which provides a more defensible basis for comparative statements under uncertainty.

Figure 4

Uncertainty and risk interval: 90% prediction interval of Top 20

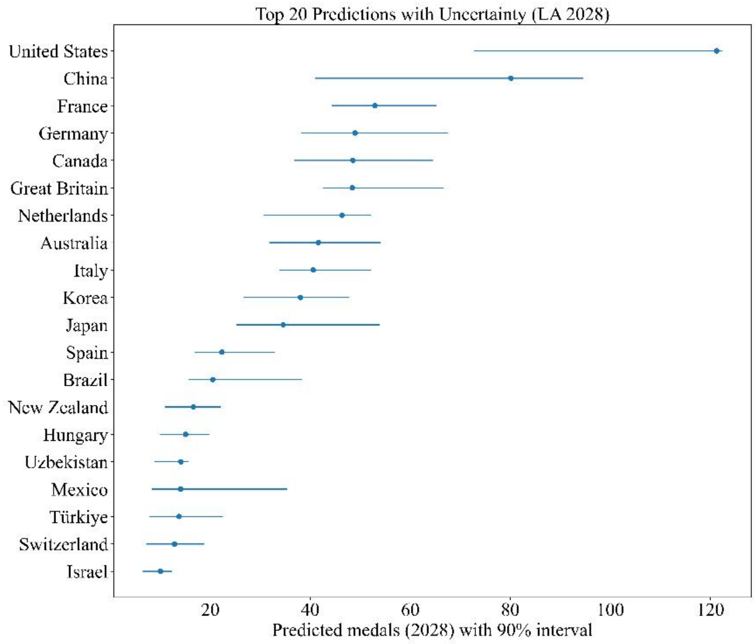


Figure 5 presents the SHAP summary plot, which provides global information on feature importance and the direction of effects. In this study, SHAP is used to decompose the selected tree-based model in an additive form, such that the prediction on the transformed scale can be written as a baseline term plus the sum of feature contributions. Because the target is log-stabilized as the natural logarithm of one plus medals, the SHAP values in Figure 5 can be interpreted as the magnitude by which a given feature increases or decreases the predicted value on the log-transformed medal scale.

Beyond a simple ranking of importance, the horizontal location of each point indicates effect direction. Two variables, athlete count and event count, display the clearest monotonic patterns: high values, shown in warm colors, are concentrated on the positive SHAP side, whereas low values, shown in cool colors, cluster on the negative side. This indicates that, conditional on the other covariates, larger delegations and broader coverage of medal events lead the model to predict higher medal totals. The underlying mechanism is straightforward. Medal production is constrained by the opportunity set: sending more athletes and entering more events implies more starts and more chances to reach the podium, which, in expectation, increases total medals. In this sense, athlete count and event count function primarily as proxies for participation opportunities, and their SHAP advantages reflect a structural relationship between opportunity constraints and the attainable upper bound of medals rather than a direct short-run causal impact of any single policy intervention.

A second layer of drivers is associated with gross domestic product and lagged total medals. Higher gross domestic product tends to correspond to more positive SHAP values, suggesting that countries with larger economic scale are more likely to sustain higher medal production, plausibly through broader training infrastructure, talent development pipelines, scientific support, and capacity to participate in international competition. The positive contribution of lagged medals highlights persistence and path dependence: historically strong teams are more likely to maintain stable coaching systems, sport-specific traditions, and talent supply, which translates into stronger performance in subsequent editions. Importantly, these factors typically operate jointly with participation proxies, and their interpretation is closer to long-run capability and resource endowment than to a single lever that can be adjusted in the short term.

It is also notable that population size and gross domestic product per capita appear lower in the SHAP ranking and do not exhibit a uniform direction. This is consistent with intuition. A larger population constitutes a potential talent pool, but conversion into medals depends on institutional capacity and sport structure. Moreover, when gross domestic product, gross domestic product per capita, and participation measures are correlated, tree models may distribute explanatory credit across correlated predictors, making any single variable appear less dominant in the SHAP ordering. Similarly, the appearance of a partially negative contribution for sport count is more plausibly attributable to correlation and substitution among predictors. Once event count already captures the breadth of event-level coverage, sport count may carry additional information about dispersion rather than intensity, and its marginal contribution can attenuate or even reverse sign. This should not be interpreted as evidence that participating in more sports reduces medals. Rather, it suggests that, after controlling for event coverage and delegation size, the remaining incremental information conveyed by sport count is weak and may point toward a more diffuse participation structure. Dependence plots are therefore necessary to verify whether this pattern is robust.

Finally, the host indicator ranking lower in SHAP does not imply the absence of a host advantage. A more plausible interpretation is that much of the host-related advantage is absorbed by observable channels already included in the feature set, such as delegation size, event coverage, prior performance, and macroeconomic resources. In addition, host effects are likely heterogeneous across countries and editions, depending on investment intensity, strategic emphasis, and contemporaneous changes in the event program. For this reason, host effects are more appropriately examined through an event-study difference-in-differences design that targets dynamic changes around hosting, rather than expecting SHAP within a predictive model to deliver a stable estimate of an average host increment. This division of labor closes the loop of "prediction, interpretation, and quasi-causal diagnosis": SHAP explains the association structure learned by the model, while the event-study design is used to probe the dynamic increment attributable to the hosting event.

Figure 5

SHAP analysis diagram

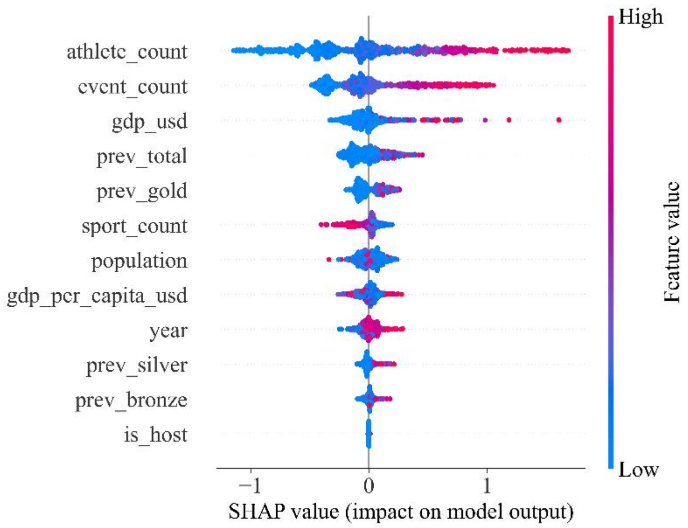


Figure 6a indicates an overall positive and nonlinear relationship between previous-edition total medals and the corresponding SHAP values. As previous total medals increase from very low levels to the mid-range, the SHAP values move from negative to positive and rise rapidly. At higher levels of previous total medals, the

contribution remains positive but shows clear diminishing returns and a tendency to level off. This pattern suggests that, conditional on the other covariates, the model treats prior performance as a key prior for subsequent medal production, consistent with pronounced path dependence. Historically strong teams are more likely to sustain training systems, sport-specific traditions, coaching structures, and talent pipelines, thereby maintaining higher medal outputs in later editions. More importantly, the color gradient reveals stratification by athlete count at similar levels of previous total medals. Observations with larger delegations more frequently appear in higher SHAP regions, implying a synergistic effect between historical strength and participation intensity. Mechanistically, previous total medals capture the stock of existing competitive capability, whereas athlete count is closer to an intensity measure that reflects how strongly this capability is translated into competition opportunities. The combination of these two components increases the likelihood that latent advantages are converted into realized medals.

Figure 6b shows the clearest structure between athlete count and SHAP values. The relationship is strongly positive and largely monotonic, with evident threshold and saturation features. In the low athlete-count range, SHAP values are predominantly negative, indicating that small delegations are associated with lower predicted values on the log-transformed medal scale. As athlete count increases to a moderate range, SHAP values quickly become positive and continue to rise. At higher delegation sizes, the contribution remains positive but the slope becomes flatter, reflecting diminishing marginal returns to further expansion. This shape aligns closely with plausible medal-generation mechanisms. Expanding delegation size initially increases starts, advancement opportunities, and the probability of reaching finals, which raises expected medal totals. Once coverage approaches saturation, additional athletes are more likely to be allocated to events with stronger competition or to weaker disciplines, thereby yielding smaller marginal gains. The color encoding of previous bronze medals suggests further heterogeneity at similar athlete-count levels. Higher previous bronze counts are associated with more positive SHAP values in parts of the distribution, implying that an established breadth of competitive capacity, approximated here by bronze accumulation, may improve the conversion efficiency from scale investment to medal outcomes. In other words, delegation size is not purely a quantity effect; its payoff depends on sport foundations and depth of the talent pool, with bronze medals serving as an imperfect but informative signal of that depth.

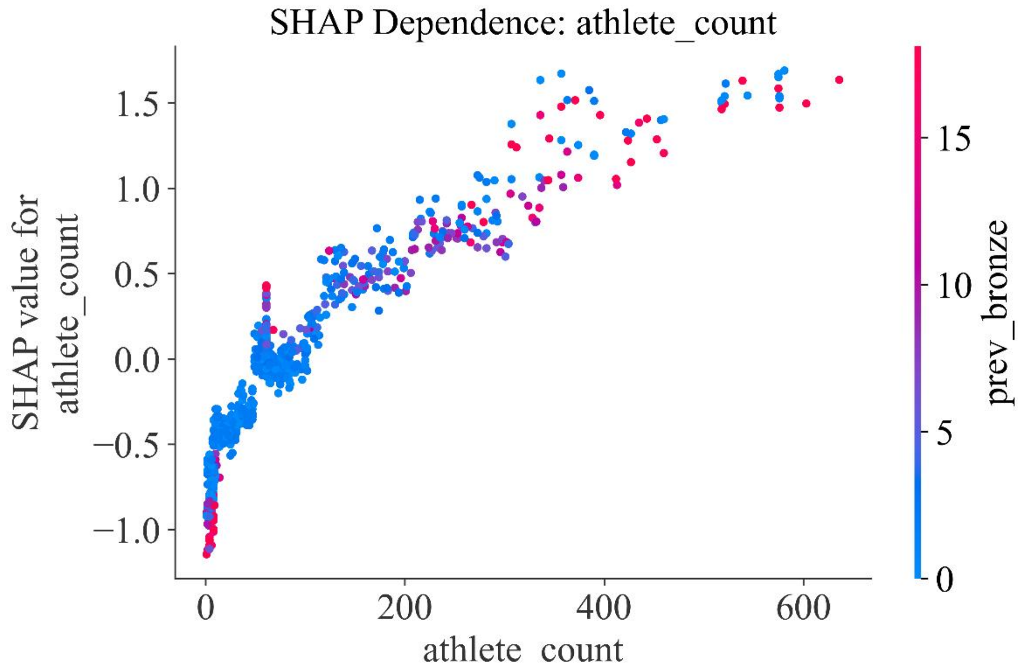
Figure 6c shows that event count is also positively associated with SHAP values, although the pattern appears more segmented and discretized than for athlete count. This is expected because the event-count variable takes relatively discrete values, producing visible vertical banding. In the low event-count range, SHAP values are mostly negative. As event coverage increases to moderate and higher levels, SHAP values turn positive and rise, indicating that broader event-level coverage is treated by the model as an important driver of higher predicted medal totals. From a mechanism perspective, event count represents opportunity space rather than direct athletic capability. Entering more events expands the number of entry points into medal contention and increases the number of potential podium outcomes, which naturally pushes total medals upward. The color gradient for gross domestic product per capita further highlights heterogeneity in conversion efficiency. At comparable levels of event coverage, higher gross domestic product per capita is more often associated with higher positive SHAP values, suggesting that resource and institutional conditions may strengthen the ability to translate breadth into podium outcomes. Such conditions plausibly include higher-quality training facilities, sport science support, professionalized program management, and accumulated international competition experience. Put differently, event count captures breadth of participation, whereas gross domestic product per capita reflects the efficiency with which breadth is converted into medals.

Figure 6

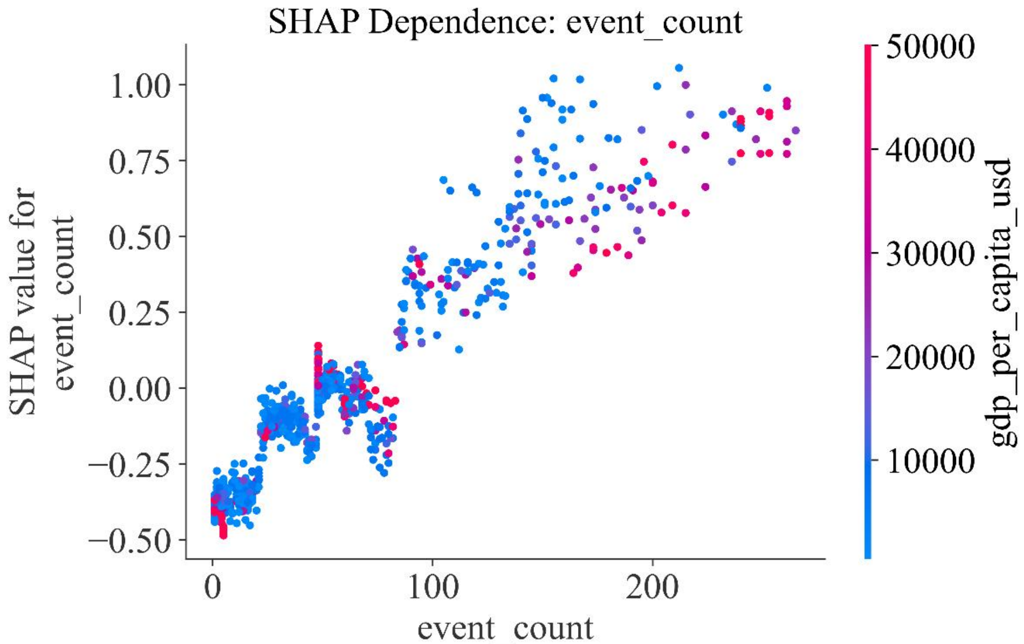
SHAP dependence plots for key drivers of LA 2028 medal predictions

a. prev_total (colored by athlete_count)

Insights of Economics System
WISDOM ACADEMIC ISSN: 3080-731X | E-ISSN: 3080-7328 Volume 2 , Issue 1



b. athlete_count (colored by prev_bronze)



c. SHAP dependence plot for event participation breadth (event_count), colored by GDP per capita (gdp_per_capita_usd).

Insights of Economics System
ISSN: 3080-731X | E-ISSN: 3080-7328 Volume 2 , Issue 1

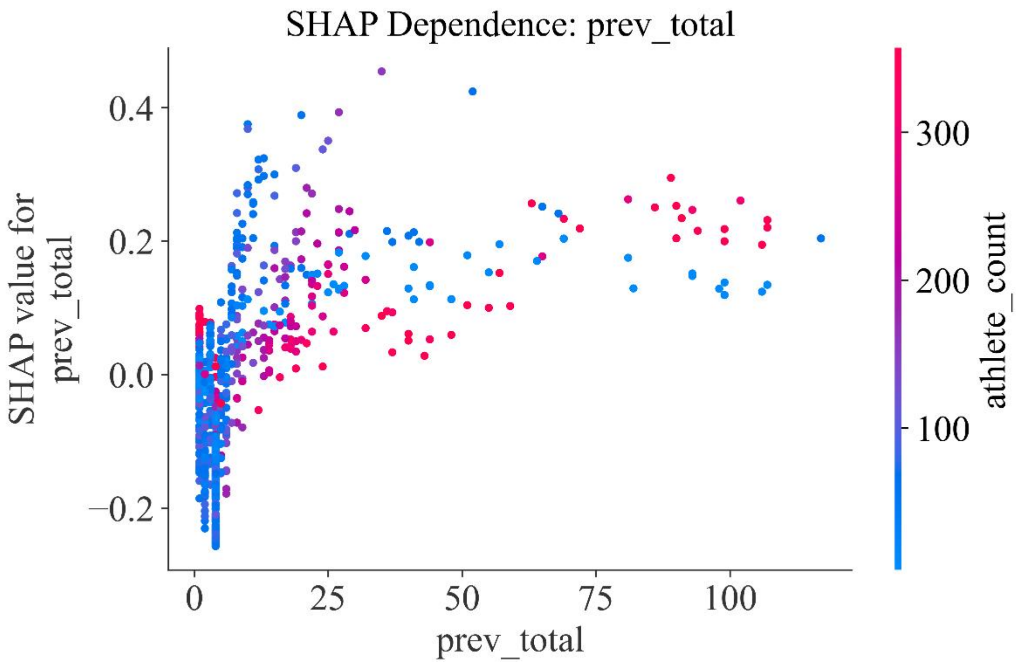
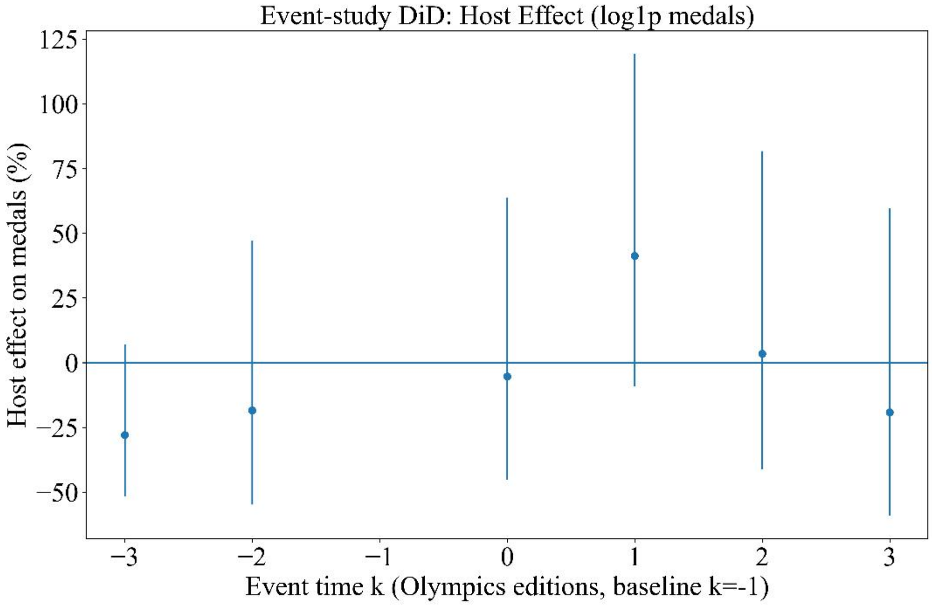


Figure 7 presents the dynamic host-country effect estimated using an event-study difference-in-differences specification, with event time minus one serving as the reference period. The estimates exhibit generally wide confidence intervals that frequently include zero, and the coefficients around the host edition and adjacent editions are especially uncertain. Under the current sample size and the set of available controls, the empirical evidence for a statistically precise and stable magnitude of the host advantage is therefore limited. The results are more appropriately interpreted as suggestive directional signals rather than as a tightly identified average treatment effect. In addition, the pre-host coefficients are not uniformly close to zero, which raises concerns about whether the parallel-trends assumption can be fully satisfied or convincingly assessed within a short event-time window and a limited number of host observations. When considered alongside the forecasting results, a coherent interpretation emerges: the machine learning models primarily capture predictable variation transmitted through observable channels of capacity and scale, whereas causal identification of host effects is constrained by the sparsity of host-country events, substantial cross-country heterogeneity, and potential omitted factors such as changes in event programs, delegation selection policies, and time-varying investment intensity. Future work that incorporates more granular sport-level or event-level data and a longer temporal window may improve the precision and credibility of host-effect estimates.

Figure 7

Dynamic analysis diagram of event study DiD

Insights of Economics System
WISDOM ACADEMIC ISSN: 3080-731X | E-ISSN: 3080-7328 Volume 2 , Issue 1



5. Conclusion

This study proposes an integrated, reproducible workflow for Olympic medal forecasting that links (i) year-grouped machine-learning prediction, (ii) bootstrap-based uncertainty quantification, (iii) SHAP-driven interpretability at both global and country levels, and (iv) an event-study DiD design aimed at separating host-related dynamics from long-run national strength.

Empirically, out-of-fold results suggest that the models capture the dominant cross-country structure under a realistic “predict-by-edition” constraint, while prediction variance increases materially among top medal nations—consistent with stronger exposure to unobserved factors such as program structure, delegation composition, and cycle-specific shocks.

In the LA 2028 application, the framework produces coherent head-to-mid-tier stratification (USA ~121; China ~80; followed by a compact second tier), and the bootstrap intervals highlight that uncertainty remains substantial even where point forecasts are high.

From an interpretation standpoint, SHAP results consistently emphasize that participation scale and event participation breadth (athlete_count and event_count) are the most robust positive contributors, reflecting an “opportunity set” mechanism: larger delegations and broader event coverage increase the number of medal-contesting chances and thus raise expected totals. Macro scale (GDP) and performance inertia (prev_total) contribute in a complementary way, capturing long-run resource and system persistence. Importantly, the relatively lower marginal contribution of the host dummy within the predictive model should not be over-interpreted as “no host advantage”; rather, it suggests that much of what is observable about hosting may already be absorbed by delegation scale, coverage, and prior strength, and that any remaining host effect is likely heterogeneous across countries and cycles.

Finally, the event-study DiD analysis offers a cautious message: under available controls and the finite set of hosts, dynamic host-effect estimates exhibit wide confidence intervals and do not yield a stable, precisely identified average treatment path, reinforcing the need for richer covariates (e.g., sport/event-level structure, qualification quotas, investment proxies) and robustness checks tailored to staggered adoption and treatment heterogeneity. In practice, this implies that policy-facing interpretation should prioritize interval-aware comparisons (overlapping ranges) and mechanism-consistent narratives (opportunity/scale vs. resources/inertia), while treating “host advantage” as a separate, data-demanding identification problem rather than a byproduct of a prediction model.

Author contributions: Conceptualization, J.Z. and L.P.; methodology, L.P.X; software, L.P.; validation, J.Z. and L.P.; formal analysis, L.P.; investigation, J.Z.; resources, J.Z.; data curation, J.Z. writing—original draft preparation, J.Z and L.P.; writing—review and editing, J.Z and L.P.; visualization, J.Z and L.P.; supervision, J.Z and L.P.; project administration, J.Z; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

Funding: This research received no external funding.

References

[1] Balmer, N. J., Nevill, A. M., & Williams, A. M. (2003). Modelling home advantage in the Summer Olympic Games. *Journal of Sports Sciences, 21*(6), 469–478. https://doi.org/10.1080/0264041031000101890

[2] Bernard, A. B., & Busse, M. R. (2004). Who wins the Olympic Games: Economic resources and medal totals. *Review of Economics and Statistics, 86*(1), 413–417. https://doi.org/10.1162/003465304774201824

[3] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[4] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

[5] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

[6] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[7] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

[8] Grimes, A. R., Kelly, W. J., & Rubin, P. H. (1974). A socioeconomic model of national Olympic performance. *Social Science Quarterly, 55*(4), 777–783.

[9] Johnson, D. K. N., & Ali, A. (2004). A tale of two seasons: Participation and medal counts at the Summer and Winter Olympic Games. *Social Science Quarterly, 85*(4), 974–993.

[10] Lozano, S., Villa, G., Guerrero, F., & Cortés, P. (2002). Measuring the performance of nations at the Summer Olympics using data envelopment analysis. *European Journal of Operational Research, 148*(2), 308–325.

[11] Lui, H. K., & Suen, W. (2008). Men, money, and medals: An econometric analysis of the Olympic Games. *Pacific Economic Review, 13*(1), 1–16. https://doi.org/10.1111/j.1468-0106.2007.00386.x

[12] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing*

Systems 30 (NeurIPS 2017). https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

[14] Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games* (Vol. II, pp. 307–317). Princeton University Press.

[15] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006