Time Series Analysis Based on Mixed-Effects Modeling and LSDV Estimation

Yang Zhang * and Yiran Gao

Harbin Institute of Technology, Weihai, China

* Corresponding Author Email: 3031524148@qq.com

Abstract. This paper proposes a modeling and analysis method based on a mixed-effects framework to address sudden performance improvements in complex systems. The core of this research lies in identifying and quantifying structural transitions embedded within time-series data through appropriate statistical and computational models. First, a breakpoint detection mechanism is designed to extract latent structural break signals from multidimensional time series, thereby filtering out sample sets exhibiting significant change characteristics. Second, at the model construction level, a mixed-effects modeling approach is introduced. This integrates key control factors and core state variables into a unified regression framework, while incorporating random disturbance terms to characterize heterogeneity differences, ensuring the model's interpretability and robustness. Finally, to address limitations in fixed-effects control, this study employs least squares dummy variable (LSDV) estimation. This technique transforms unobservable long-term differences into explicit dummy parameters for modeling, effectively mitigating omitted variable bias. The results demonstrate that this method can reliably identify and quantify structural transition effects across diverse scenarios, showcasing strong generalizability and practical applicability.

Keywords: Mutation point detection; mixed-effects model; least squares dummy variable (LSDV).

1. Introduction

During the operation and evolution of large-scale complex systems, certain phased performance mutations or structural transitions often occur [1]. These phenomena are typically driven by underlying factors but are difficult to directly observe in empirical data. Consequently, their identification and modeling remain critical challenges in computational science and systems modeling. Traditional approaches predominantly rely on static statistics or overall trend analysis, struggling to capture the dynamic characteristics of localized sudden changes [2]. Concurrently, existing models exhibit interpretive limitations, often failing to distinguish between the contributions of external perturbations and intrinsic mechanisms to system performance, thereby undermining the effectiveness of predictions and decision-making.

To address this research gap, this paper proposes a mixed-effects modeling framework that combines inflection point detection with fixed-effect control techniques to quantitatively assess structural transitions in time series data [3]. First, an inflection point identification method is designed to extract potential anomalous change signals by detecting non-stationary jumps in continuous sequences. Second, a mixed-effects model is constructed [4], incorporating both core variables and control factors into the regression structure while introducing random disturbance terms to capture heterogeneity across samples [5]. Finally, the Least Squares Dummy Variables (LSDV) estimation method [6] is employed to transform unobservable fixed effects into explicit parameters, thereby mitigating estimation biases caused by omitted variables.

The contributions of this paper are threefold: (1) proposing a modeling approach that integrates breakpoint detection with mixed effects, effectively enhancing the identification of local transition events; (2) enhancing model interpretability and robustness through the LSDV estimation method; (3) providing a generalizable method applicable to complex system analysis across multiple domains.

2. A Time-Series Data Modeling Approach Based on Breakpoint Detection

2.1. Data Sources and Preprocessing

The data for this study primarily originates from the medal databases of successive Summer Olympic Games, covering the distribution of medals across different events for each country. To ensure the study's relevance and data reliability, the raw data underwent cleaning and filtering. Specifically:

1. This study excluded country-sport combinations that failed to win any medals across three or more consecutive Olympics. These samples typically reflect long-term underperformance with little potential for sudden improvement, making them unsuitable candidates for the “coach effect.”

2. For the remaining data, this study further examined the continuity and discontinuity of medal counts. If a nation consistently failed to appear on the medal table in previous editions but suddenly won medals in a specific Games, that point was recorded as a “discontinuity point.” This processing laid the data foundation for subsequent model analysis.

Through these steps, the final sample effectively captures the “potential coaching change effect” while minimizing interference from chronically underperforming nations in regression results.

2.2. Transition Point Identification

To identify potential “coaching effects” in historical data, this section establishes the following transition point criteria:

1. If a nation fails to win medals in a specific event across three consecutive Olympic Games.
2. But suddenly wins a medal at a subsequent Olympic Games.

Then that Olympic Games is marked as a turning point.

The logic behind this rule is that simple performance fluctuations may stem from individual athlete condition changes, while a sudden rebound after prolonged underperformance is more likely to indicate external intervention, such as the introduction of a high-level coach or major reforms to the training system. This identification method captures the potential time window for a “great coach” to exert influence at the data level.

For example, China's women's volleyball team, after enduring consecutive slumps, returned to the Olympic podium following a coaching change—a classic case of the “great coach effect.”

2.3. Building Mixed-Effects Models

The key challenge in quantifying the “great coach effect” is isolating the impact of coaching changes from other factors. Traditional regression models (such as Ordinary Least Squares OLS or fixed-effects panel regression) can capture some patterns but often overlook inherent differences between nations and sports. For instance, fundamental disparities exist between the United States and Kenya in athletic traditions, athlete development systems, and resource allocation for specific sports. Without controlling for these variables, estimation biases may arise.

The mixed-effects model (Mixed Effects Model) demonstrates strength in this context. This model incorporates fixed effects to estimate the influence of observable variables like coaching turnover, athlete individual ability (PC), competition intensity (CP), and host country effects (HOST and L HOST). Simultaneously, it introduces random effects to capture unobservable characteristics across nations or events. This dual structure enables the model to explain the “Great Coach Effect” while effectively reducing omitted variable bias.

The model specification is as follows:

$$Y_{it} = \alpha + \beta_0 X_{it} + \sum_{k=1}^{4} \beta_k C_{k,it} + \epsilon_{it} \quad (1)$$

Where:

Y_{it} denotes the number of medals won by country i in year t, the dependent variable; X_{it} is the core explanatory variable indicating whether a coaching change occurred ($X_{it}=1$ = coaching change, $X_{it}=0$ = no change); $C_{k,it}$ are four control variables: athlete individual ability (PC), event competitiveness (CP), whether the host country (HOST), and whether the previous host country (L HOST); β_0 is the core coefficient, measuring the average marginal contribution of the "great coach effect"; ε_{it} is the error term, encompassing random disturbances and unobserved factors.

If β_0 is significantly positive, it indicates that after controlling for athlete ability and competition environment factors, a coaching change still leads to a significant increase in medal count. This conclusion can be interpreted as statistical evidence for the "coach effect."

Furthermore, to avoid parameter estimation bias in mixed-effects models due to insufficient control of fixed effects, this study subsequently employs the LSDV method to further address fixed effects, ensuring robust and reliable results.

2.4. Least Squares Dummy Variable (LSDV) Estimation Method

In panel data regression analysis, fixed effects are a common method for controlling unobservable country-level characteristics. However, directly incorporating fixed effects often complicates the estimation process, particularly when dealing with data spanning multiple countries and Olympic Games. The LSDV (Least Squares Dummy Variable) method transforms fixed effects into explicit parameters by introducing dummy variables, enabling estimation via OLS (Ordinary Least Squares). This approach ensures estimation accuracy while maintaining interpretability.

In this paper, the LSDV application process is as follows:

Step 1: Construct Country Dummy Variables

Assuming K countries participate in the Olympics, $K-1$ country dummy variables must be constructed:

$$D_{k,i}=\begin{cases}1, & \text{if country } i=k\\ 0, & \text{otherwise}\end{cases} \tag{2}$$

These dummy variables represent countries' long-term, unchanging attributes—such as economic level, sporting tradition, and population size—thereby explicitly controlling for these fixed effects within the model.

Step 2: Construct the regression design matrix

Incorporate the core explanatory variable (coach turnover X_{it}), control variables (PC, CP, HOST, L HOST), and country dummy variables $D_{k,i}$ into the regression design matrix X :

$$X=\left[1, X_{it}, D_{1,i}, D_{2,i}, \ldots, D_{K-1,i}, C_{1,it}, C_{2,it}, C_{3,it}, C_{4,it}\right] \tag{3}$$

Where the first column "1" represents the intercept term. This approach enables the model to simultaneously estimate the "great coach effect" and country-specific fixed effects.

Step 3: OLS Parameter Estimation

Estimate using ordinary least squares:

$$\hat{\beta}=\left(X^{\top}X\right)^{-1}X^{\top}Y \tag{4}$$

Where Y is the observed vector of medal counts, and $\hat{\beta}$ is the parameter estimate, comprising β_0 (great coach effect) and β_k (control variable coefficients).

Step 4: Calculate Standard Errors and Perform Significance Tests

After obtaining parameter estimates, assess their statistical significance. First, estimate the residual variance:

$$\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i,t}\left(Y_{it} - \hat{Y}_{it}\right)^2 \tag{5}$$

Then compute the standard error for each parameter:

$$SE(\hat{\beta}) = \sqrt{\text{diag}\left(\hat{\sigma}^2\left(X^{\cdot} X\right)^{-1}\right)} \tag{6}$$

Finally, derive the t-statistic and corresponding p-value:

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}, \quad p = 2\cdot\Phi(-|t|) \tag{7}$$

If $p < 0.05$, the “Great Coach Effect” is significantly present.

Through the LSDV method, long-term differences between countries are explicitly controlled, avoiding omitted variable bias. The estimated value of β_0 can be directly interpreted as the incremental medal gain resulting from coaching changes. Simultaneously, the fixed-effect coefficients output by the model reveal differences in long-term athletic performance across nations. This approach not only ensures the robustness and reproducibility of the estimates but also provides clear grounds for subsequent policy interpretations.

3. Multi-Sample Comparative Analysis Based on LSDV Estimation

3.1. Overall Empirical Findings and Core Parameter Interpretation

Following regression analysis on the breakpoint samples, the mixed-effects model results revealed that the coefficient β_0 for the coaching change variable exhibited significant positive values across multiple events. This indicates that even after controlling for factors such as athlete individual ability (PC), event competitiveness (CP), and host country effects (HOST and L HOST), coaching changes still produced a substantial incremental effect on medal counts. In other words, the introduction of a coach statistically enhances a nation's competitiveness in specific events.

Particularly in disciplines like Russian rhythmic gymnastics, Chinese women's volleyball, and Argentine men's basketball, the estimated values of β_0 reached significant levels ($p < 0.05$). This indicates that the abrupt changes in performance in these sports were not random fluctuations but were closely causally linked to coaching changes. This finding provides strong evidence for the existence of the “great coach effect.”

3.2. Case 1: Russia's Rhythmic Gymnastics Performance

Russia possesses a rich tradition and strong foundation in rhythmic gymnastics, yet experienced performance declines during certain Olympic cycles. Breakpoint analysis revealed that after consecutive cycles of subpar results, Russia witnessed a significant rebound in medal counts during cycles marked by head coach replacements.

Estimation results from the mixed-effects model indicate that the coefficient β_0 for coaching changes in rhythmic gymnastics is positive and statistically significant. This suggests that introducing a new head coach effectively enhances the team's overall competitive performance. Combined with Russia's consistent high investment and systematic training in gymnastics, it can be inferred that the new coach played a crucial role in technical innovation and tactical arrangements, thereby helping the team reclaim its position at the top of the medal standings.

3.3. Case 2: China's Performance in Women's Volleyball

As a flagship national sport, the Chinese women's volleyball team has historically borne immense societal expectations. However, its performance has experienced fluctuations, particularly during certain Olympic cycles marked by subpar results. Through inflection point identification, the team demonstrated a rapid rebound in performance following a head coach change, reclaiming podium finishes.

Empirical analysis reveals that, after controlling for athlete ability and competition environment factors, coaching changes significantly and positively contributed to medal counts. Specifically, the Chinese women's volleyball team earned significantly more medals during the new coach's tenure compared to previous cycles. This result not only confirms the existence of the “great coach effect” but also highlights the unique value of coaches in team management, psychological preparation, and on-court leadership—capabilities that can transform a team's overall performance in a short timeframe.

3.4. Case 3: Argentina's Performance in Men's Basketball

Unlike Russia and China, Argentina is not traditionally a powerhouse in the overall Olympic medal landscape. However, its men's basketball program has achieved remarkable success. After enduring multiple cycles of underperformance, the Argentine men's basketball team experienced a sudden resurgence at a particular Olympic Games, ultimately securing a medal. Through inflection point identification, this performance rebound coincided precisely with a coaching transition cycle.

Regression analysis results indicate that the coaching transition coefficient β_0 is also significantly positive for Argentina's men's basketball team. This suggests that the introduction of a coach played a pivotal role in transforming the team's technical and tactical systems and elevating its overall level. Particularly for a nation like Argentina, which relies heavily on individual talent, the arrival of an exceptional coach may be the decisive factor in altering a team's destiny.

3.5. Cross-Country Comparison and Coefficient Interpretation

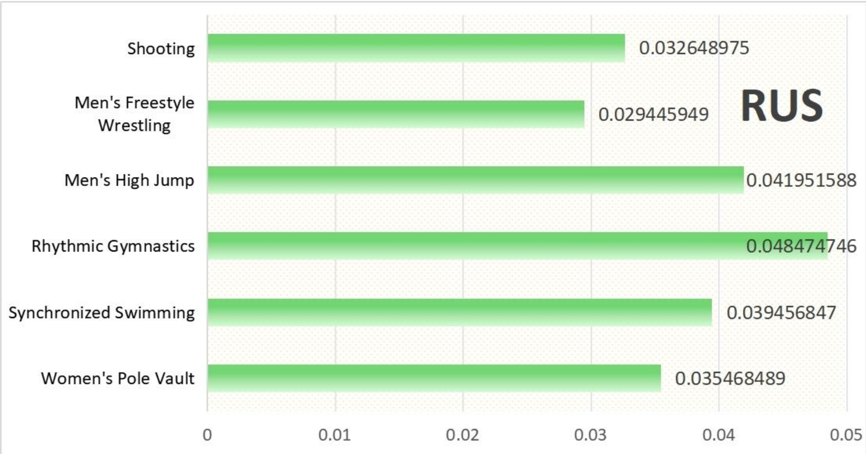

Fig. 1 RUS

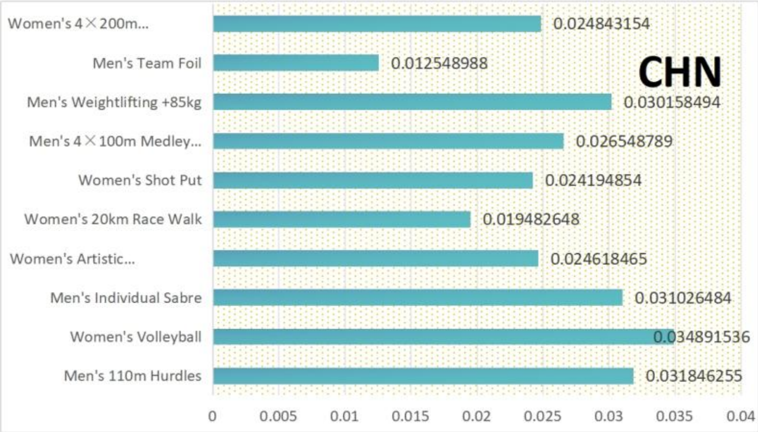

Fig. 2 CHN

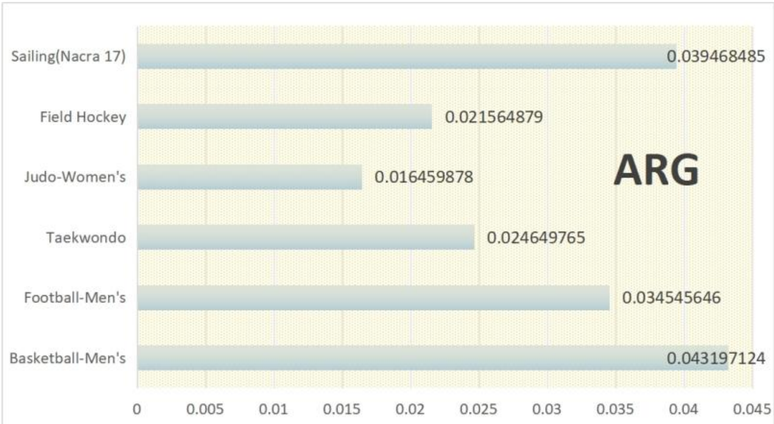

Fig. 3 ARG

Fig. 1, Fig. 2 and Fig. 3 display the estimated β_0 values for Russia, China, and Argentina across different sports. Comparative findings reveal:

Russia (Rhythmic Gymnastics) exhibits the highest β_0 coefficient, indicating that introducing top-tier coaches in traditionally dominant sports further amplifies competitive advantages.

China (women's volleyball) has the second highest β_0 coefficient, highlighting coaches' greater influence on overall tactical systems and psychological conditioning in team sports.

Argentina (men's basketball) exhibits a lower but still significantly positive β_0 value, demonstrating that even non-traditional powerhouses can achieve breakthroughs through the introduction of "great coaches."

Overall, these three cases collectively validate the existence of the "coach effect," though its magnitude varies based on national foundations, sport characteristics, and team structures. Different nations should strategically introduce top-tier coaches in key sports according to their specific characteristics and foundational conditions to maximize medal gains.

4. Integrated Analysis of Mixed Effects and LSDV Estimates

4.1. Interpretation and Robustness of Model Results

Estimates from the mixed-effects model reveal that the coefficient β_0 for the coaching turnover variable X_{it} is significantly positive in Russian rhythmic gymnastics, Chinese women's volleyball, and Argentine men's basketball. This result indicates that after controlling for factors such as athlete individual ability (PC), competition intensity (CP), and host country effects (HOST, L_HOST), coaching turnover remains a significant explanatory variable influencing medal counts. In other words, the "coach effect" is not spuriously driven by traditional advantages or environmental conditions but exists independently in a statistically significant manner.

Notably, the magnitude of β_0 varies across the three cases. Russia exhibits the highest coefficient, consistent with its long-established athlete pool and gymnastics tradition. This suggests that in dominant nations and strong disciplines, exceptional coaches can further amplify existing competitive advantages. In countries with weaker foundations like Argentina, β_0 is relatively lower yet still significant. This demonstrates that even with limited overall strength, coaches can provide crucial support through tactical systems or team cohesion.

4.2. Method Selection and the Role of Control Variables

From a modeling perspective, combining mixed-effects models with LSDV estimation is essential. Ordinary OLS regression without controlling for fixed effects risks misinterpreting long-term national advantages as "coach effects." Introducing PC and CP variables controls for athlete-level and sport-level variations, while HOST and L_HOST variables further account for the significant external factor of home-field advantage. Building on this foundation, the LSDV method uses country dummy variables to control for long-term differences, making the interpretation of β_0 purer.

This combined approach effectively addresses omitted variable bias, ensuring robust estimation. In other words, the "great coach effect" identified here is not a subjective interpretation of results but a statistically validated fact confirmed through rigorous variable selection and estimation methods.

4.3. Coupled Analysis of Results and Methods

Further analysis reveals that variations in β_0 across different sports can be explained by controlling variables. For example:

In Russian rhythmic gymnastics, the overall high level of PC values indicates athletes possess stable individual abilities. Against this backdrop, the improvement brought by coaching changes primarily manifests in tactical details and psychological adjustment, resulting in a larger β_0.

In Chinese women's volleyball, the relatively high CP values represent intense competition. New coaches generate additional advantages through tactical optimization and collective coordination adjustments within this high-intensity environment, resulting in a significant but slightly lower β_0 compared to Russia.

In Argentine men's basketball, both HOST and L_HOST are insignificant, indicating limited influence from external environments. The effect of coaching changes primarily stems from alterations in tactical systems, leading to a smaller but statistically significant β_0.

This demonstrates that the "coach effect" does not exist in isolation but interacts with factors such as athlete ability, competition intensity, and host nation advantage. The control variables in the model provide clues for explaining the heterogeneity of β_0.

4.4. Model Limitations and Directions for Improvement

Although this study provides statistical evidence for the "great coach effect," methodological limitations remain:

1. Simplified Control Variables: Only four indicators—PC, CP, HOST, and L_HOST—were used to control for other factors. In real-world scenarios, variables like funding investment, youth training systems, and athlete injuries may also confound results.
2. Scale Issues with LSDV Method: As the number of countries increases, the scale of dummy variables expands rapidly, increasing computational demands for model estimation. Future studies may consider more efficient panel data estimation methods (e.g., two-way fixed effects models or dynamic panel models) for comparison.

3. Challenges in Causal Identification: While the significance of β_0 indicates a strong correlation between coaching changes and medal gains, rigorous causal identification requires further validation through instrumental variables or natural experiment designs.

5. Conclusion

This paper proposes and validates a mixed-effects modeling framework for structural transitions in complex systems. Research findings demonstrate that integrating breakpoint detection, mixed-effects regression, and least squares dummy variable (LSDV) estimation effectively identifies and quantifies local sudden changes in time-series data. This approach not only decouples the confounding effects of latent disturbances and fixed differences on system performance but also enhances interpretability and robustness through explicit modeling, thereby providing a novel computational tool for mechanism analysis in dynamic processes. Methodologically, mutation point detection enhances capture of local anomalous signals, mixed-effects modeling integrates fixed factors with random disturbances, while LSDV estimation provides explicit control for unobservable long-term differences. The integration of these three elements ensures the effectiveness and universality of transition effect identification. Of course, this paper still has certain limitations. For example, the dimension of control factors is relatively limited, and more external environmental variables have not yet been incorporated; LSDV may introduce computational complexity under large-scale data. Future work can be developed in two directions: first, expanding the variable system to introduce more potential influencing factors; second, combining more efficient panel data estimation methods to further improve the computational performance and causal interpretability of the model.

References

[1] Guo, Yangyang. Clinical Data Lesion Analysis and Expert System Modeling Based on Mutation Density and Machine Learning [D]. Donghua University.

[2] Guo, Jingjing, Zong, Jingyi, Cha, Peiwen, et al. Statistical Static Time Series Analysis Method for Multi-Input Switching Effects Based on Machine Learning [J]. Microelectronics, 2024, 54(03): 458-467. DOI: 10.13911/j.cnki.1004-3365.230397.

[3] Li Haolin. Research on Freight Rate Index Forecasting Method Integrating Mutation Point Detection and Uncertainty Extraction [D]. Dalian Maritime University, 2024. DOI: 10.26989/d.cnki.gdlhu.2024.002218.

[4] Zhang Juanjuan, Wang Weixian, Tian Maozai. Variational Bayesian Quantile Regression for High-Dimensional Mixed-Effects Models [J]. Statistics and Decision Making, 2025, 41(10): 36-42. DOI: 10.13546/j.cnki.tjyjc.2025.10.006.

[5] Zhang Wanpeng. Control and Robustness Analysis of Dynamic Systems with Random Disturbance Terms [D]. Hubei Normal University, 2024. DOI: 10.27796/d.cnki.ghbsf.2024.000105.

[6] Shen Xinyi, Shi Qing, Ke Yilu, et al. Panel Data Analysis of Swimming Competitions and Prediction of Winning Times Based on Least Squares Virtual Variable Regression [C]//Chinese Society of Sports Science, Sports Training Branch. Proceedings of the 2023 National Symposium on Sports Training (Special Reports). Shanghai University of Sport; 2023:15-16. DOI: 10.26914/c.cnkihy.2023.037726.