

Scenario-Based Salary Allocation in Major League Baseball under Fixed Budget Constraints

Jun Lin^{1,*}, Yuchen Li²

¹*International Department, The High School Affiliated to Renmin University of China, Beijing, China*

²*Zhenhai High School, Ningbo, China*

**Corresponding Author. Email: seirei28@126.com*

Abstract: This paper studies how Major League Baseball teams should allocate payroll under fixed budget constraints. Rather than focusing only on the direct correlation between salary inequality and team performance, the paper links salary structure, team total utility, real payroll, and winning percentage together to analyze the relatively optimal solution. Using MLB team-year data from 1985 to 2016, the study constructs player-level utility from offensive and pitching contributions, combines it to the team level, and calculates salary structure with the Gini coefficient. Then it estimates the team production relationship and compares different payroll structures among different budget groups using scenario-based analysis. The results indicate that team total utility is the strongest predictor of winning percentage. Also, the immediate impact of salary inequality becomes weaker compared to the total utility. Indeed, real payroll is shown to be positively related to performance. These scenario-based findings further suggest that no single salary structure is the most optimal: concentrated structures work rather well in lower budget settings, whilst moderately concentrated structures perform better in medium and high-budget groups. Overall, the paper focuses on a more practical comparison of salary allocation strategies under fixed budget constraints.

Keywords: Major League Baseball; salary structure; team utility; fixed budget constraints

1. Introduction

1.1. Research background and problem statement

In professional team sports, payroll is important to teams and clubs, but the size of payroll alone does not ensure performance. It depends on the allocation of resources among players, which means teams with similar spending levels may achieve very different results. Studies of Major League Baseball and related professional sports have suggested that total payroll and salary distribution share some correlation with team performance [1–3]. As such, the salary structure is a significant issue for the examination of team performance. Major League Baseball (MLB) provides a beneficial context in which to investigate this question, since various teams have vastly inconsistent payroll scale and internal salary distribution. Some teams have relatively equal salary spread across the roster, while

others concentrate a large share of salary on a smaller number of star players. Research in MLB has evaluated directly if wage disparity or salary dispersion is related to team productivity and performance [4–7]. The primary question of this paper consequently, is which salary structure performance is more efficient with a fixed budget. More exactly, the following research questions regarding the relationship among payroll concentration, team productive strength and winning percentage, and it will investigate whether the salary structure effect is determined by budgetary level.

1.2. Research objectives and contribution

The research paper has two broad aims. Firstly, it empirically evaluates the relation between team performance, salary inequality, team total utility, and real payroll at the team-year scale. Second, it employs these empirical findings to explore differences in salary structures in different budgets. This paper comes double as a contribution. From a logical perspective, it makes a connection between the salary system, and winning percentage (as illustrated by team total utility) instead of just holding payroll inequality as a separate explanatory variable. In practice, it shifts from precise value estimation to scenario-based comparison by comparing the various forms of payroll structures under low-, medium-, and high-budget conditions. This is an empirical analysis, using team and player data from the Lahman Baseball Database [8].

2. Literature review and research positioning

2.1. Empirical evidence from existing studies

Evidence regarding pay inequality in MLB and team performance is mixed. Some studies find that salary dispersion is negatively related to team productivity or performance, suggesting that excessive inequality may create coordination costs or reduce team cohesion [4, 5, 9]. Other studies suggest that the relationship can be more complicated, depending on the measurement of salary dispersion, the role of star players, and the way team performance is modeled [6, 7, 10].

Research on payroll and performance also shows that total spending matters, but it does not fully explain team success. Studies of MLB payroll have found relationships between payroll and winning, but the the relationship may vary in different empirical specifications [1–3]. Other studies further exams the the distribution of salary within teams and their total performance, showing that payroll allocation can also mater besides the effect of overall payroll level [11, 12].

But in practice, these results are usually affected by many factors, such as model specification, sample choosing, and variable definition. Specifically, studies vary with respect to how they measure salary inequality, how they control total payroll, and how they model relationship between salary allocation and team performance [10, 13, 14]. As a result, it is difficult to draw a simple conclusion about which is the most optimal salary structure [10].

2.2. Research gap and positioning of this study

This study differs from previous studies in two ways [10]. First, rather than directly exam the effect of salary inequality on winning percentage, an indicator, total utility, is created. This enables the analysis to separate players' contribution to the team from salary distribution.

Second, rather than searching for one optimal level of salary inequality, this paper creates a framework based on different scenario that relates to different real-budget groups. This approach is more practical to MLB teams, which usually have to allocate payroll under a given budget constraint instead of calculating a specific optimal level of inequality [2,3]. Overall, this paper provides a practical

insight to front office of each teams by offering a decision-make framework in order to maximize the team performance as well as improving financial capability. Broader theoretical and empirical discussions of pay dispersion and team performance have also appeared in related literature [15–17].

3. Data sources and variable construction

3.1. Data sources and sample description

This study uses Major League Baseball (MLB) data from 1985 to 2016. This sample region matches the coverage of the official salary records used in the study and provides sufficient evidence for team-level analysis across different competitive environments.

The final empirical dataset is constructed at the team-season level. However, several variables are first built at the player-team-season level and then aggregated to the team level. The main datasets are drawn from the Lahman Baseball Database, including *People*, *Salaries*, *Teams*, *Pitching*, and *Batting*.

The data construction has four steps. First, batting and pitching statistics are used to construct player-level contribution measures. Second, salary information is used to calculate within-team payroll inequality. Third, team records are used to calculate winning percentage. Fourth, total team payroll is converted into CPI-adjusted real payroll in constant 2016 dollars in order to avoid the effect of inflation.

Players who appeared for more than one team in the same season are calculated separately at the player-team-season level so that both contribution and salary are allocated to the correct team.

3.2. Measurement of individual player utility

To connect salary allocation with team performance, this study constructs an individual player utility measure based on offensive and pitching contribution.

3.2.1. Offensive contribution: Runs Created (RC)

Offensive contribution is measured by Runs Created (RC), defined as:

$$RC = H + 2B + 2(3B) + 3(HR),$$

where H denotes hits, $2B$ doubles, $3B$ triples, and HR home runs. This measure gives greater weight to extra-base hits and serves as a simple proxy for offensive production.

3.2.2. Pitching contribution: Runs Saved Above Average (RSAA)

Pitching contribution is measured by Runs Saved Above Average (RSAA). Since pitching data are recorded in outs, innings pitched, individual ERA, and league-average ERA are calculated as:

$$IP = \frac{IP_{outs}}{3}, \quad ERA = \frac{ER \times 9}{IP}, \quad lgERA = \frac{\sum ER_{league}}{\sum IP_{league}} \times 9.$$

Here, ER denotes earned runs allowed, and $lgERA$ is calculated separately for each $(yearID, lgID)$ combination using aggregate league totals. Based on these quantities, RSAA is defined as:

$$RSAA = (lgERA - ERA) \times \frac{IP}{9}.$$

A positive RSAA indicates above-average pitching performance, while a negative RSAA indicates below-average performance. For players with $IP = 0$, RSAA is set equal to 0.

3.2.3. Standardization and construction of total player utility

Because RC and RSAA are measured on different scales, both variables are standardized before being combined. Let zRC_i and $zRSAA_i$ denote the standardized offensive and pitching contributions of player i , respectively. Individual utility and team total utility are then defined as:

$$U_i = zRC_i + zRSAA_i, \quad TU_j = \sum_{i=1}^{n_j} U_i.$$

Here, TU_j denotes team total utility, and n_j is the number of player-team observations for team j in a given season.

3.3. Measurement of team salary structure

The main variable in this study is salary inequality within teams. It is measured by the Gini coefficient, which is calculated for each team-year observation using player salary records from the *Salaries* table, and the formula is given by:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |S_i - S_j|}{2n^2 \bar{S}}$$

where S_i and S_j represent player salaries within the same team-year and $\bar{S}$ is the corresponding average salary. A larger Gini coefficient indicates a more concentrated payroll distribution.

Because the Gini coefficient is a relative measure of within-team salary distribution, inflation adjustment is not applied to its construction. By contrast, total team payroll is also calculated as a scale variable and converted into CPI-adjusted real payroll in constant 2016 dollars for the later inflation and budget-group analysis.

3.4. Construction of the team performance variable

Team performance is measured by winning percentage, defined as:

$$Win\% = \frac{W}{W + L}$$

where W is the number of wins and L is the number of losses in a given season. Winning percentage is used instead of total wins because it provides a standardized measure of team performance across seasons.

3.5. Construction of the final team-year panel

The final dataset is summarised by the team-year level by merging team total utility, salary Gini coefficient, winning percentage, nominal payroll, and CPI-adjusted real payroll by $(yearID, teamID)$. In this framework, the Gini coefficient represents each team's salary concentration, while real payroll indicates the inflation-adjusted scale of team financial resources.

3.6. Descriptive statistics and preliminary feature analysis

Before estimating the team production relationship, the study first summarizes the main variables and provides valid evidence on salary structure, team utility, and team performance.

Table 1. Descriptive statistics of the main variables

Statistic	Team Total Utility	Salary Gini	Winning Percentage	Real Payroll (M, 2016\$)
Mean	0.0141	0.5634	0.5001	55.34
Std. Dev.	11.2868	0.1037	0.0687	52.48
Min	-35.1165	0.0000	0.2654	0.145
25th pct.	-7.2513	0.5107	0.4506	21.54
Median	0.4947	0.5792	0.5000	40.56
75th pct.	7.4130	0.6328	0.5494	71.87
Max	30.6649	0.8026	0.7160	530.79
Observations	916	916	916	916

Table 1 shows the variation in team total utility, salary inequality, and real payroll across team-year observations. In particular, CPI-adjusted payroll differs widely across the sample, suggesting that salary structure should be evaluated together with budget level rather than in isolation.

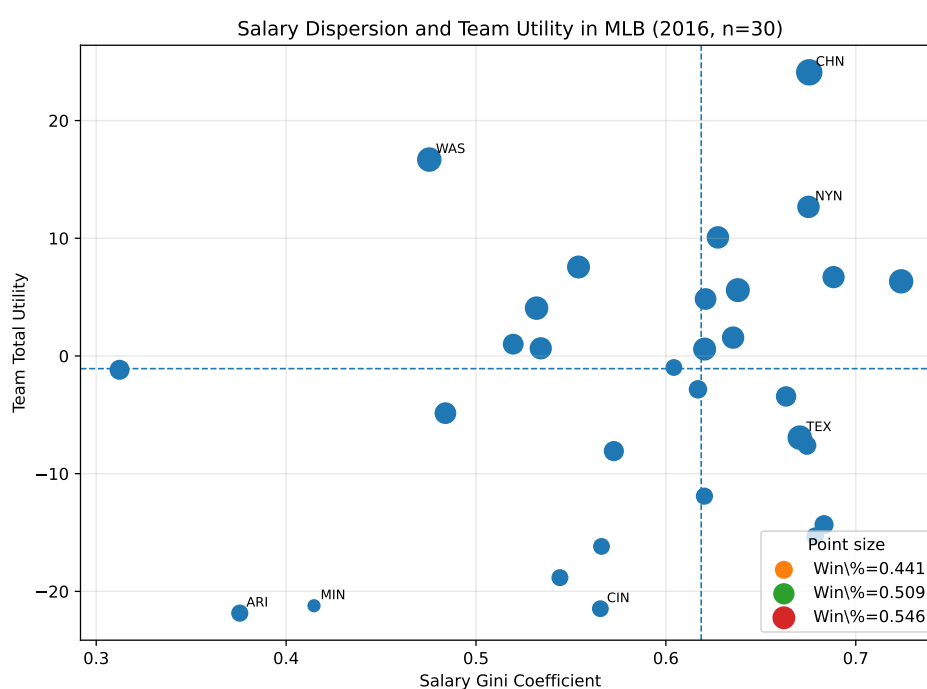


Figure 1. Salary dispersion and team utility in MLB (2016). Dashed lines indicate sample medians, and point size reflects winning percentage.

Figure 1 presents a cross-sectional situation of the 2016 season, showing how teams differed in salary dispersion, team utility, and winning percentage.

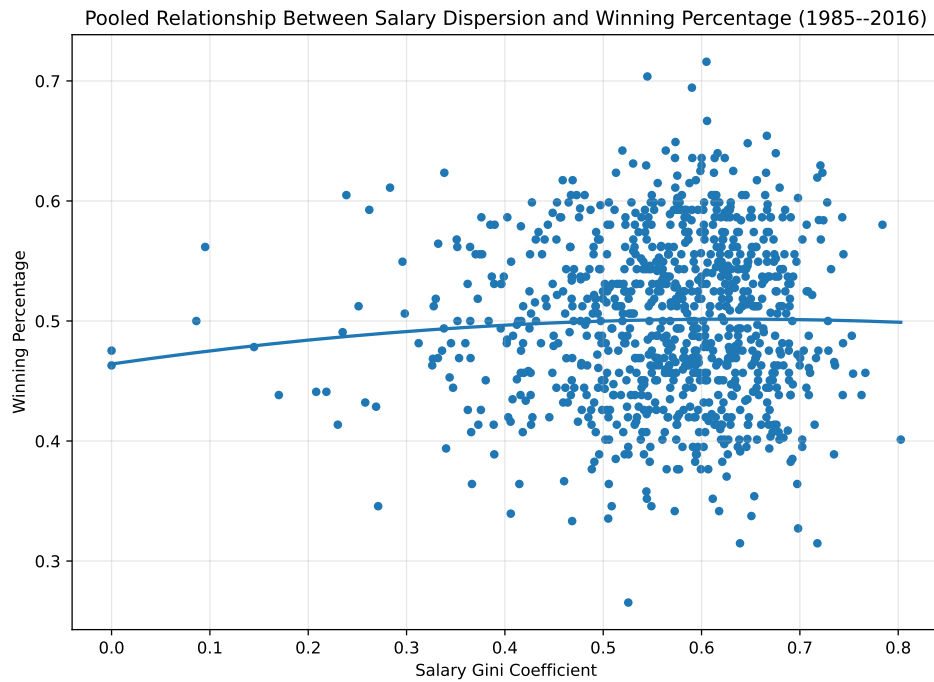


Figure 2. Pooled relationship between salary dispersion and winning percentage, 1985–2016. The quadratic fit is shown for descriptive purposes only.

Figure 2 plots the pooled relationship between salary dispersion and winning percentage over 1985–2016. The visual relationship appears weak, which is consistent with the later regression results.

4. Empirical calibration and scenario-based allocation framework

4.1. Research logic and framework

This chapter links the empirical analysis to the paper’s main decision problem: under a fixed payroll budget, what type of salary structure is more likely to support stronger team performance?

The logic of the study follows four steps. First, real payroll defines the team’s overall resource constraint. Second, different allocation choices generate different salary structures, which are reflected in the within-team Gini coefficient. Third, payroll allocation influences roster quality, summarized in this paper by team total utility. Finally, team total utility and salary structure together shape winning percentage.

Accordingly, the analytical chain of the paper can be summarized as:

$$\text{Budget} \rightarrow \text{Salary Structure} \rightarrow \text{Team Utility} \rightarrow \text{Winning Percentage}.$$

This model implies that payroll does not mechanically affect performance. A bigger budget can enhance outcomes, but only if it is converted to productive team strength. Therefore, the chapter first estimates the team production relationship, and then uses the results to help establish a scenario-specific comparative analysis of salary structures among the budget settings.

4.2. Empirical specification of the team production relationship

The paper examines the relationship between roster productivity, salary structure, and budget scale with team performance by estimating a team-level production relationship involving winning percentage as a function of team total utility, salary inequality, and real payroll.

The baseline specification is:

$$Win_{jt} = \alpha + \beta_1 TU_{jt} + \beta_2 G_{jt} + \varepsilon_{jt},$$

where Win_{jt} is the winning percentage of team j in season t , TU_{jt} is team total utility, and G_{jt} is the team-level salary Gini coefficient.

To allow for a possible nonlinear relationship between salary inequality and performance, the model is extended as:

$$Win_{jt} = \alpha + \beta_1 TU_{jt} + \beta_2 G_{jt} + \beta_3 G_{jt}^2 + \varepsilon_{jt}.$$

Because the linear and squared Gini terms may be highly correlated, the richer specifications use a centered Gini measure:

$$G_{jt}^c = G_{jt} - \bar{G},$$

where $\bar{G}$ is the sample mean of the Gini coefficient. The quadratic specification then becomes:

$$Win_{jt} = \alpha + \beta_1 TU_{jt} + \beta_2 G_{jt}^c + \beta_3 (G_{jt}^c)^2 + \varepsilon_{jt}.$$

To account for differences in budget scale across teams, real payroll is added:

$$Win_{jt} = \alpha + \beta_1 TU_{jt} + \beta_2 G_{jt}^c + \beta_3 (G_{jt}^c)^2 + \beta_4 Payroll_{jt} + \varepsilon_{jt}.$$

Finally, year fixed effects are included to absorb league-wide influences that vary over time:

$$Win_{jt} = \alpha + \beta_1 TU_{jt} + \beta_2 G_{jt}^c + \beta_3 (G_{jt}^c)^2 + \beta_4 Payroll_{jt} + \gamma_t + \varepsilon_{jt}.$$

Here, $Payroll_{jt}$ denotes CPI-adjusted real payroll in constant 2016 dollars, measured in millions for interpretability, and γ_t represents year fixed effects.

4.3. Parameter calibration

The empirical calibration is done in three steps. Model 1 contains team total utility and salary Gini coefficient. Model 2 includes centered quadratic Gini as a test for possible nonlinearity. Model 3 presents additional real payroll and year fixed effects. All the models are estimated using ordinary least squares with heteroskedasticity-robust standard errors.

The results shows a clear pattern. Throughout all of the specifications, the team total utility is positive and especially significant, suggesting that productive roster strength is the most stable predictor of winning percentage. Unlike the salary Gini coefficient which is not so robust.

Thus, the regression analysis does not provide adequate evidence of an interior optimum for salary inequality. Such findings imply that, in large part at least, salary structure matters in the extent to which it's an effective facilitator of payroll becoming team utility.

4.4. Scenario-Based allocation framework

Given the regression results, the salary allocation problem is reformulated as a scenario-based comparison under a fixed payroll budget. The team is assumed to face the following budget constraint:

$$\sum_{i=1}^n S_i = B,$$

where S_i denotes the salary of player i , and B is total team payroll.

Instead of solving a continuous optimization problem over all individual salaries, the paper adopts a practical categorical framework. Salary structures are grouped into three broad types:

1. **Balanced structure:** payroll is distributed relatively evenly across players;
2. **Moderately concentrated structure:** payroll is focused on several core players while maintaining roster depth;
3. **Highly concentrated structure:** a large share of payroll is allocated to a small number of highly paid players.

To use this framework, team-year observations are grouped by budget level and salary structure. With the bottom 1% and top 1% of real payroll entries removed, the remaining sample is separated into low, medium, and high-budget groups based on their CPI-adjusted payroll.

This produces a 3×3 scenario table. For each budget–structure combination, the study shows mean team total utility, mean winning percentage, mean salary Gini coefficient, mean real payroll, and the number of observations.

Within this standard, team total utility serves as the key intermediate indicator because it captures how effectively a salary structure converts player's salary into a quantitative variable, while winning percentage still represents the final team performance. Since the regression does not identify a reliable optimum value of salary inequality, this comparison based on scenario provides a more practical way to evaluate which combination performs best in different salary structures.

5. Empirical results and scenario-based salary structure comparison

5.1. Estimation results of the team production relationship

Table 2 reports the estimation results for the team production relationship under three specifications: a baseline linear model, a quadratic model with centered Gini terms, and a fuller model including real payroll and year fixed effects.

Across all three models, team total utility remains positive and highly significant, indicating that total team productivity is the most influential determinant of winning percentage in the sample. By contrast, the direct effect of salary inequality is much weaker. In Model 1, the salary Gini coefficient is only weakly significant, while in Model 2 and Model 3 both the centered Gini term and its squared term become statistically insignificant. This suggests that the data do not provide reliable evidence for a stable nonlinear relationship between salary inequality and team performance.

Table 2. Regression results for the team production relationship

Variables	Model 1	Model 2	Model 3
Intercept	0.488***	0.501***	0.486***
Team Total Utility	0.005***	0.005***	0.005***
Salary Gini	0.022*		
Centered Gini		0.011	0.000
Centered Gini Squared		-0.089	-0.089
Real Payroll (2016\$ Millions)			0.000**
Year Fixed Effects	No	No	Yes
Observations	916	916	916
R^2	0.606	0.606	0.624

Notes: The dependent variable is team winning percentage. All models are estimated using OLS with HC3 robust standard errors. Model 1 contains team total utility and salary Gini. Model 2 replaces raw Gini with centered Gini and its squared term. Model 3 also incorporates CPI-adjusted real payroll (in 2016 dollars) and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Model 3 also shows that real payroll has a positive and statistically significant effect on winning percentage. However, the coefficient on team total utility remains much more stable and more strongly significant than the salary structure variables. Overall, the results suggest that payroll affects performance through its ability to generate productive team strength primarily rather than through salary inequality alone.

5.2. Scenario-based results and managerial implications across budget groups

Table 3 reports the scenario-based results under fixed real-budget constraints. Team-year observations are grouped by budget level and salary structure, and each cell reports the number of observations, mean team total utility, mean winning percentage, mean salary Gini coefficient, and mean real payroll in CPI-adjusted 2016 dollars.

Table 3. Scenario-based results under fixed real-budget constraints, 1985–2016

Budget Group	Structure Group	Obs.	Mean TU	Mean Win%	Mean Gini	Mean Real Payroll (M, 2016\$)
Low Budget	Balanced	178	-0.6791	0.4910	0.4366	18.21
Low Budget	Moderate	76	-1.8796	0.4956	0.5756	23.83
Low Budget	Concentrated	45	3.1244	0.5134	0.6514	25.65
Medium Budget	Balanced	72	-0.4631	0.4997	0.4947	55.03
Medium Budget	Moderate	124	0.6376	0.5070	0.5810	54.78
Medium Budget	Concentrated	102	-2.7882	0.4857	0.6640	57.29
High Budget	Balanced	49	-0.8600	0.4945	0.4985	151.18
High Budget	Moderate	98	2.1524	0.5101	0.5817	147.99
High Budget	Concentrated	152	0.5944	0.5053	0.6643	143.20

Two main patterns appear. First, the effect of salary structure varies across budget groups. In the low-budget group, concentrated structures perform best, with the highest mean winning percentage

and team total utility. In the medium- and high-budget groups, moderately concentrated structures perform best, while concentrated structures perform especially poor in the medium-budget group.

Second, no single salary structure dominates across all budget settings. This suggests that payroll design should be evaluated in context of budget level rather than against one universal basis. For management, the implication is that the most effective salary structure is the one that converts available payroll into the greatest productive team strength within a given budget environment.

6. Conclusion

The paper analyses how MLB teams can allocate payroll in the situation of a fixed budget constraint. Instead of making a decision based on a static explanatory variable like salary inequality alone, the research builds on a theory-based approach connecting budget, salary structure, team total utility, and winning percentage.

The empirical findings reveal that team total utility is the strongest, most stable predictor of performance. On the other hand, the direct nonlinear impact of salary inequality is not measured in accurate form which is sufficient to verify a sole optimal Gini coefficient. In other words, the payroll structure should not be judged based solely on a single indicator as a universal benchmark.

The scenario-based comparison provides a more practical conclusion. Under low-budget conditions, concentrated salary structures are associated with the highest average winning percentage, while under medium- and high-budget conditions, moderately concentrated structures perform best. These findings suggest that the most effective payroll structure depends on budget level and that front-office decisions should focus on converting available payroll into productive roster strength rather than just pursuing equality or concentration in the abstract.

Further research will be committed to grow the sample, to include a deeper analysis of player contribution, and to investigate effects of trades, injuries and roster adjustment on payroll use further. Nevertheless, the paper demonstrates that the allocation of salaries is best viewed as a problem of strategy based on budget, instead of the one-size-fits-all pursuit of optimal inequality.

References

- [1] Wiseman, F., & Chatterjee, S. (2003). Team payroll and team performance in Major League Baseball: 1985–2002. *Economics Bulletin*, 1(2), 1–10.
- [2] Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). Testing causality between team performance and payroll: The cases of Major League Baseball and English soccer. *Journal of Sports Economics*, 3(2), 149–168.
- [3] Veach, T. (2010). Competitive balance and team payroll: The case of Major League Baseball. *Major Themes in Economics*, 12(1), 65–86.
- [4] Depken, C. A. (2000). Wage disparity and team productivity: Evidence from Major League Baseball. *Economics Letters*, 67(1), 87–92.
- [5] Annala, C. N., & Winfree, J. A. (2011). Salary distribution and team performance in Major League Baseball. *Sport Management Review*, 14(2), 167–175.
- [6] Tao, H., Chuang, H.-H., & Lin, E. S. (2016). Compensation and performance in Major League Baseball: Evidence from salary dispersion and team performance. *International Review of Economics & Finance*, 43, 151–159.
- [7] Vyskocil, S. (2019). Salary Dispersion and Team Performance in Major League Baseball. *Explorations in Economic and Business History*, 12(1), Article 13.
- [8] Lahman, S. (2025). *Lahman Baseball Database*. Society for American Baseball Research (SABR).
- [9] Breunig, R., Garrett-Rumba, B., Jardin, M., & Rocaboy, Y. (2014). Wage dispersion and team performance: A theoretical model and evidence from baseball. *Applied Economics*, 46(19), 2275–2294.
- [10] Treber, J., Levy, M., & Vickers, A. (2024). Wage dispersion and team performance: A review of 25 years of research. *American Behavioral Scientist*, 68(10), 1240–1263.
- [11] Jewell, R. T., & Molina, D. J. (2004). Productive efficiency and salary distribution: The case of U.S. Major League Baseball. *Scottish Journal of Political Economy*, 51(1), 127–142.

- [12] Lu, X., & Weinberg, B. A. (2018). Team payroll, pitcher and hitter payrolls and team performance. *Contemporary Economic Policy*, 36(2), 301–315.
- [13] Frick, B., Prinz, J., & Winkelmann, K. (2003). Pay inequalities and team performance: Empirical evidence from the North American major leagues. *International Journal of Manpower*, 24(4), 472–488.
- [14] Kamada, T., & Okumura, M. (2014). Team performance and within-team salary disparity. *Economics Bulletin*, 34(3), 1514–1525.
- [15] Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5), 841–864.
- [16] Levine, D. I. (1991). Cohesiveness, productivity, and wage dispersion. *Journal of Economic Behavior & Organization*, 15(2), 237–255.
- [17] Katayama, H., & Nuch, H. (2011). A game-level analysis of salary dispersion and team performance in the National Basketball Association. *Applied Economics*, 43(10), 1193–1207.