Better or Worse?

Biased Decisions from Human-Machine Collaboration

Byunghun Hahn*

October 1, 2025

Abstract

Korea Baseball Organization (KBO) introduced the replay review system to improve the accuracy and fairness of the umpire decisions in 2014. However, we show that the winning probability of the home team has significantly increased after the implementation, which implies that the introduction of the replay review system deteriorated the fairness of the umpire's decisions. Our interpretation is that it is because of the imperfect system design. Because the replay review chances are limited to two times, the umpires ended up having incentives to give favorable decisions for the home team.

*Department of Economics, Seoul National University

1 Introduction

Humans have long relied on machines to lighten the physical burden. But future advances will focus on collaboration between humans and machines that share decision-making authority. Such changes are expected to increase the reliability and accuracy of the decision-making process, but successful cooperation is not guaranteed all the time.

There is a line of research with implications that go beyond the sports industry (Abramitzky et al., 2012; Anbarci et al., 2016; Mills, 2017; Paulsen, 2025). Our research likewise has implications for human–machine collaboration.

A recent development in baseball games gives us a unique opportunity to observe the cooperation between humans and machines in decision making. In a baseball game, umpires possess absolute authority. They make various decisions on almost every aspect of a game and their decisions cannot be questioned.

On 22 July 2014, the Korea Baseball Organization (KBO) introduced replay review system (RRS from now on). Under this new system, each team has two chances per game to challenge the umpire's decision. When a challenge is made, umpires should review the video records of the play. This system was adopted abruptly during the season, after a series of controversial judgments provoked a fan to assault the umpire.

We empirically show that RRS resulted in a counterintuitive effect. Home team winning percentage significantly rose by 5%p. We show this using a difference-in-differences specification.

Studies have shown that there are various types of bias in umpire decisions, such as home advantage (Hsu, 2024; Smith and Groetzinger, 2010) and racial discrimination (Parsons et al., 2011; Pope et al., 2018; Price and Wolfers, 2010) and nationalistic bias (Thrane, 2025). However, only few studies have examined how monitoring systems affect umpire bias, and their findings are straightforward: such systems increase fairness (Dufner et al., 2023; Holder et al., 2022; Parsons et al., 2011). To the best of our knowledge, we are the first to examine the interaction between RRS and home team advantage in baseball, and our results suggest that RRS increased the home advantage. We claim that this is because the opportunity to review decisions is limited to two times per team and the umpire's individual incentive works against the intention of introducing RRS.

2 Data and descriptive statistics

We collect match-level data for all KBO regular-season games (2009-2024) from NAVER.com, South Korea's most used portal site, and the KBO Official Yearbook. For each game, we collect the date, home and away teams, final score, umpires and attendance.

Our main outcome is the home team's winning probability. Figure 1 plots the home winning percentage and average attendance for each half-season.[1] A notable pattern is that the home winning percentage rose sharply after RRS was introduced in mid-2014 and remained elevated until 2019. Interestingly, this pattern disappeared during COVID-19, when the home team's winning percentage returned to pre-2014 levels. These shifts challenge the common belief that RRS would lead to fairer decisions and reduce home team advantage. Instead, it appears to have increased home advantage, with the magnitude of the impact affected by attendance levels.

Table 1 summarizes institutional changes over time. The number of teams increased from 8 to 10 by 2015, raising the number of games from 512 to 704 per season.

Table 2 provides summary statistics. Our dataset includes 10,213 games, with an average home winning percentage of 51.4%. The average attendance and attendance rate are 9.8k and 48%, respectively.

[1]Each season is split into two halves, separated by the All-Star Game break in mid-July. The only exception is 2020, when the All-Star Game was rescheduled for September due to COVID.

3 Econometric analyses and results

We adopt a difference-in-differences strategy to estimate the effect of RRS on home team winning percentage. A key challenge is the absence of a control group, as the system was implemented league-wide simultaneously.

To address this, we exploit variation in game attendance by dividing each season's games into high- and low-attendance groups based on the yearly median. Games below the median serve as the control group, while those above serve as the treated group. The underlying idea is that if crowds amplify the psychological pressure on umpires under RRS, the impact should be stronger in games with higher attendance.

This approach is supported by both prior literature and our own evidence. Parsons et al. (2011) provide evidence that referees in baseball are influenced by crowd-induced social pressure. Under RRS in baseball, such pressure may be even stronger, as umpires must re-watch plays on-site while fans simultaneously view the same footage on large screens and through live broadcasts. This public scrutiny likely increases pressure, especially in crowded stadiums.

Empirically, using match-level data, we regress the home team win indicator on attendance, separately for three periods: 2009–mid-2014 (pre-review), mid-2014–2019 (post-review & pre-COVID), and 2020–2024 (post-COVID). In line with Mills and Fort (2014), we find no systematic effect of attendance on home team win probability in general. However, Table A1 shows that significant positive effects emerge only during the post-review & pre-COVID period. We interpret this as evidence that the introduction of RRS created a new channel through which crowd size affected outcomes. While attendance alone does not influence home winning probability, under RRS greater crowd pressure affected umpire behavior and thereby increased the home team winning percentage.

We use a difference-in-differences estimation to estimate the effects of replay review with the following specification:

$$\begin{aligned} I\{HomeWin\}_{ithau} = \alpha &+ \beta_1 Treat_i + \beta_2 Post_t + \beta_3 (Treat_i \times Post_t) \\ &+ \delta_t + H_h + A_a + U_u + \varepsilon_{ithau} \end{aligned} \tag{1}$$

$I\{HomeWin\}_{ithau}$ is a binary indicator equals 1 if the home team wins match i at time t, with home team h, away team a, and umpire(s) u. The fixed effects for year, home team, away team, and umpire are denoted by δ_t, H_h, A_a, and U_u, respectively.

Table 3 summarizes our key empirical results, which examine the effect of RRS on home team winning percentage. We report results from three specifications with varying fixed effects. Across all models, the DID coefficients are approximately 0.055, indicating a 5.5%p increase in home team win rates after the introduction of RRS. Appendix Table A2 shows consistent results when using attendance rate to define treatment status.

Our main empirical finding is that home team winning percentage increased following the introduction of RRS, with low-attendance games serving as a control group.

As a robustness check, we propose an alternative identification strategy using the unique timing of the implementation. RRS was introduced abruptly in the middle of the 2014 season. Since teams in the first half of that season played without RRS, and without any expectation that it would be introduced, we use first-half outcomes to predict second-half results, using five statistical learning methods. These predicted outcomes serve as the control group, while the actual second-half results constitute the treated group.

This approach assumes that team strength remains stable between the first and second halves. While this assumption is reasonable and widely accepted, we empirically validate it. Using data from all seasons between 2009 and 2024, we train prediction models on first-half games and test them on second-half outcomes.[2]

As shown in Figure 2, prediction accuracy drops sharply in 2014 across all models, then returns to prior levels in 2015. We interpret this as evidence of a structural break in the second half of 2014, caused by the introduction of replay review.

Table 4 presents the results using predicted outcomes as the control group. The estimated DID coefficients are positive and statistically significant across all five models. In the first column, which employs ridge for prediction, the DID coefficient is 0.16, approximately three times larger than in our analysis in Table 3. This is natural, as the control group in the first approach was in fact treated (but "less" treated).

In summary, both Table 3 and Table 4 provide consistent evidence that home team winning percentage increased after the introduction of RRS. Table 3 suggests that this effect is stronger in high-attendance games, while Table 4 offers additional support using statistical learning-based predictions as the control group.

[2]Inputs include the two teams, match date, holiday indicator, and doubleheader status.

4 Theory

In this section, we show that RRS can deteriorate fairness. We will define the following parameters.

- $x \in \{0, 1\}$: umpire's judgement, where $x = 0$ is a judgement favorable to the away team, and $x = 1$ is a judgement favorable to the home team.
- $\theta \in [0, 1]$: a signal in each situation that the umpire receives, representing the probability that a decision favorable to the home team is correct.
- $\mu \in \{0, 1\}$: a signal from RRS, which is perfectly accurate.

Under this setting, the umpire's decision rule can be represented by the function $f(\theta)$:

$$f(\theta) = \begin{cases} 1 & \text{if } \theta \geq \frac{1}{2}, \\ 0 & \text{if } \theta < \frac{1}{2}. \end{cases}$$

We will assume that overturning a judgement against the home team leads to a greater loss in the umpire's utility.

- $\alpha \in [0, 1]$: probability that the home team asks for RRS.
- $\beta \in [0, 1]$: probability that the away team asks for RRS.
- $-\pi$: umpire's utility when a judgement unfavorable to the home team is overturned.
- $-\varphi$: umpire's utility when a judgement unfavorable to the away team is overturned.
- $\pi > \varphi$
- If the judgement is upheld, umpire utility is 1.

When an umpire receives a signal θ and makes a favorable judgement for the home team ($x = 1$), the umpire's utility function is:

$$u(\theta, x = 1) = ((1 - \beta) + \beta\theta) \cdot 1 - \beta(1 - \theta)\varphi \tag{2}$$

Similarly, when the umpire makes a judgement unfavorable to the home team ($x = 0$), we can write:

$$u\left(\theta, x=0\right) = \left(\left(1-\alpha\right) + \alpha(1-\theta)\right) \cdot 1 - \alpha\theta\pi \tag{3}$$

Therefore, the umpire makes a judgment in favor of the home team ($x = 1$) if and only if:

$$(1-\beta) + \beta\theta - \beta(1-\theta)\varphi \; \geq \; (1-\alpha) + \alpha(1-\theta) - \alpha\theta\pi \tag{4}$$

Rearranging (4) yields:

$$\theta \geq \frac{\beta(1+\varphi)}{\beta(1+\varphi) + \alpha(1+\pi)} \tag{5}$$

We define θ^* as the threshold signal at which an umpire is indifferent between $x = 0$ and $x = 1$.

$$\theta^* = \frac{\beta(1+\varphi)}{\beta(1+\varphi) + \alpha(1+\pi)} < 1 \tag{6}$$

Furthermore, under the assumption of asymmetric utility loss ($\pi > \varphi$), and if α and β are of similar magnitude, then:

$$\beta(1+\varphi) < \alpha(1+\pi) \tag{7}$$

Consequently, from (5) and (7), we get:

$$\theta^* < \frac{1}{2} \tag{8}$$

Therefore, under RRS umpires now decide $f\left(\theta\right) = 1$ if $\theta \geq \theta^*$ and $f\left(\theta\right) = 0$ if $\theta < \theta^*$.

When the signal lies in the range $\theta^* < \theta < \frac{1}{2}$, an umpire, who would have chosen $x = 0$ without RRS, chooses $x = 1$ with RRS. This systematic shift biases judgements in favor of the home team.

Theorem 1. *In binary decisions under uncertainty, introducing a highly accurate but limited verification mechanism, such as RRS, can distort judgement. When incorrect decisions carry asymmetric utility losses depending on which side is harmed, decision-makers will shift their threshold to avoid the more costly type of error.*

While each team is allowed only two RRS requests per game, umpires do not know in advance when these will be used. This uncertainty forces umpires to act as if every decision could be reviewed with strictly positive probability.

Let N be the number of times in a game where $\theta^* < \theta < \frac{1}{2}$. Although the away team can challenge twice, at least $N-2$ such decisions cannot be reversed and thus will favor the home team.

Our model predicts that RRS, though intended to enhance fairness, can create a net advantage for home teams. While it corrects some errors, more judgements are biased in favor of the home team and remain unchallenged.

This aligns with evidence from KBO, where home winning percentage increased after RRS was introduced. However, during the COVID-19 pandemic, when games were played without spectators, this pattern disappeared. It can be shown that when $\pi = \varphi$, the threshold restores to $\theta^* = \frac{1}{2}$.

5 Conclusion

This study empirically demonstrates that the introduction of RRS led to an increase in home team winning percentage, unlike the results in Dufner et al. (2023); Holder et al. (2022), where VAR in soccer decreased the home advantage. We interpret this as the difference between the two systems, where the challenge opportunity is limited to two times in Korean baseball, while there is no limit in soccer.

A key implication of this study is that, under an imperfect system design, the introduction of machines may paradoxically reduce overall decision-making accuracy rather than improve it. As human–machine collaboration becomes more prevalent in the future, similar distortions in decision-making could emerge.

Figure 1: Half-season trends of home winning percentage & average attendance, 2009-2024

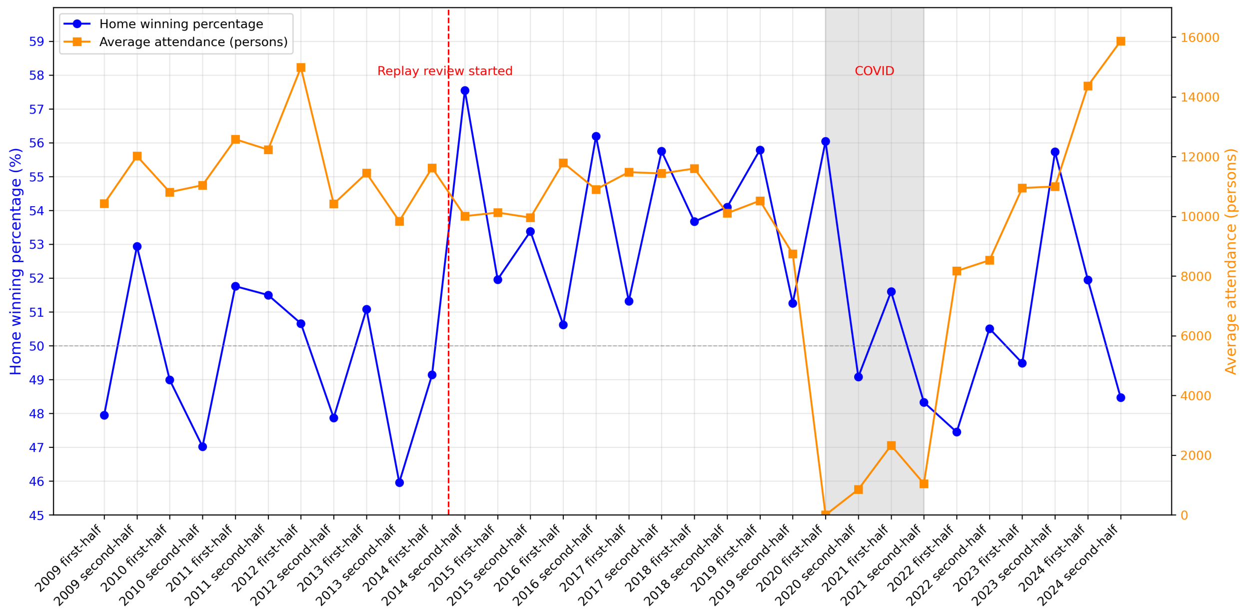
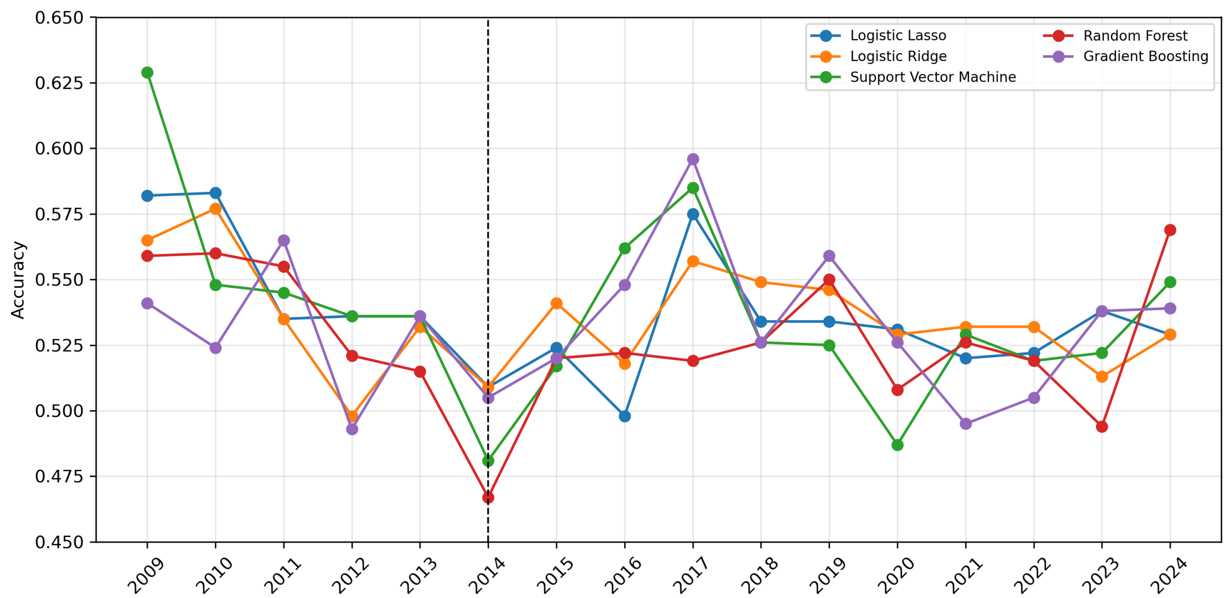


Figure 2: Half-Season Prediction Accuracy, 2009–2024



Notes: This figure shows the prediction accuracies of five statistical learning methods. Each KBO season is split into two halves by the All-Star Game; the figure reports accuracies obtained by training on the first-half games and testing on the second-half games.

Table 1: Summary statistics of KBO seasons (2009–2024).

Year	Teams	Games	Home Wins (Winning %)	Average Attendance Rate	N of Umpires	Replay Reviews	Home Challenges (Success Rate)	Away Challenges (Success Rate)
2009	8	512	254 (49.6%)	50.6%	24	–	–	–
2010	8	513	248 (48.3%)	53.6%	28	–	–	–
2011	8	513	265 (51.7%)	61.4%	26	–	–	–
2012	8	513	254 (49.5%)	65.4%	28	–	–	–
2013	9	560	274 (48.9%)	55.0%	31	–	–	–
2014	9	560	293 (52.3%)	57.3%	32	105	55 (45.5%)	50 (28.0%)
2015	10	704	370 (52.6%)	55.3%	30	407	213 (41.8%)	194 (36.6%)
2016	10	704	373 (53.0%)	56.6%	31	693	319 (31.3%)	374 (34.0%)
2017	10	704	374 (53.1%)	56.5%	33	692	359 (33.4%)	333 (27.9%)
2018	10	704	379 (53.8%)	54.6%	33	748	372 (31.7%)	376 (27.4%)
2019	10	704	382 (54.3%)	46.7%	34	787	397 (28.2%)	390 (24.4%)
2020	10	704	368 (52.3%)	2.2%	38	782	391 (28.6%)	391 (26.3%)
2021	10	705	353 (50.1%)	8.3%	39	757	351 (28.8%)	406 (24.6%)
2022	10	705	343 (48.7%)	40.0%	36	848	425 (26.4%)	423 (25.3%)
2023	10	704	368 (52.3%)	54.4%	36	978	482 (24.7%)	496 (27.0%)
2024	10	705	356 (50.5%)	74.9%	38	940	487 (24.4%)	453 (26.0%)

Table 2: Summary statistics of game-level data (2009–2024).

	N	Mean	Std. Deviation	Min	25th Percentile	Median	75th Percentile	Max
Home win indicator	10,213	0.514	0.500	0	0	1	1	1
Home Score−Away Score	10,213	-0.003	4.896	-24	-3	1	3	21
Attendance (k)	10,213	9.807	6.748	0	5	8.5	13.4	31
Attendance Rate (%)	10,213	48.746	29.644	0	26.6	46.4	71.6	100

Notes: Summary statistics are based on all games from 2009 to 2024. Games between Doosan and LG, who share the same stadium, are excluded (approximately 8–10 games per season).

Table 3: Effect of RRS on home winning probability: DID estimates using low attendance as the control group.

Independent Variables	**(1) Home Win 2009–2019**	**(2) Home Win 2009–2019**	**(3) Home Win 2009–2019**
Treat (Attendance $\geq$ Median)	0.002	0.001	0.002
	(0.021)	(0.021)	(0.021)
Post (Replay Review)	0.064	0.067	0.056
	(0.043)	(0.042)	(0.042)
Post $\times$ Treat	0.056*	0.054*	0.055*
	(0.028)	(0.027)	(0.027)
Constant	0.482***	0.491***	2.392***
	(0.028)	(0.029)	(0.146)
Year Fixed Effects	Yes	Yes	Yes
Home Team Fixed Effects	Yes	Yes	Yes
Away Team Fixed Effects	No	Yes	Yes
Umpire Fixed Effects	No	No	Yes
Observations	6,691	6,691	6,691
R^2	0.010	0.019	0.025
Mean of Dependent Variable	0.489	0.489	0.489

Notes: Standard errors clustered at the team-pair level (home $\times$ away) are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Effect of RRS on home winning probability: DID estimates using predicted outcomes as the control group.

Independent Variables	**(1) Logistic Ridge**	**(2) Logistic Lasso**	**(3) SVM**	**(4) Random Forest**	**(5) Gradient Boosting**
Treat (Attendance ≥ Median)	-0.000	-0.000	0.000	0.000	0.000
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Post (Replay Review)	-0.059	-0.031	-0.049	-0.037	-0.003
	(0.038)	(0.038)	(0.038)	(0.041)	(0.041)
Treat × Post	0.160***	0.132**	0.151***	0.137**	0.099*
	(0.058)	(0.058)	(0.057)	(0.059)	(0.059)
Constant	0.340***	0.373***	0.338***	0.349***	0.316***
	(0.064)	(0.064)	(0.064)	(0.065)	(0.065)
Home Team Fixed Effects	Yes	Yes	Yes	Yes	Yes
Away Team Fixed Effects	Yes	Yes	Yes	Yes	Yes
Umpire Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,120	1,120	1,120	1,120	1,120
R^2	0.108	0.109	0.112	0.078	0.076
Mean of Dep. Variable	0.491	0.491	0.491	0.491	0.491

Notes: Standard errors clustered at the team-pair level (home × away) are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Appendix A

Table A1: Effects of attendance on home winning probability across three time periods

Independent Variables	**(1) Home Win (2009–2014.7.16.)**	**(2) Home Win (2014.7.22.–2019)**	**(3) Home Win (2020–2024)**
Attendance (k)	0.0010	0.0045***	-0.0002
	(0.0013)	(0.0015)	(0.0012)
Constant	0.484***	0.487***	0.509***
	(0.0176)	(0.0180)	(0.0122)
Observations	2,959	3,732	3,522
R^2	0.0002	0.0024	0.0000

Independent Variables	**(4) Home Win (2009–2014.7.16.)**	**(5) Home Win (2014.7.22.–2019)**	**(6) Home Win (2020–2024)**
Attendance Rate (%)	0.0003	0.0011***	-0.0003
	(0.0003)	(0.0003)	(0.0003)
Constant	0.480***	0.478***	0.517***
	(0.0221)	(0.0199)	(0.0125)
Observations	2,959	3,732	3,522
R^2	0.0002	0.0027	0.0003

Notes: Robust standard errors in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A2: Effect of RRS on home winning probability: DID estimates using low attendance rates as the control group.

Independent Variables	**(1) Home Win 2009–2019**	**(2) Home Win 2009–2019**	**(3) Home Win 2009–2019**
Treat (Attendance $\geq$ 50%)	0.009	0.006	0.008
	(0.017)	(0.017)	(0.018)
Post (Replay Review)	0.069	0.071	0.061
	(0.046)	(0.045)	(0.045)
Post $\times$ Treat	0.041*	0.041*	0.040
	(0.023)	(0.023)	(0.024)
Constant	0.481***	0.491***	2.382***
	(0.027)	(0.024)	(0.151)
Year Fixed Effects	Yes	Yes	Yes
Home Team Fixed Effects	Yes	Yes	Yes
Away Team Fixed Effects	No	Yes	Yes
Umpire Fixed Effects	No	No	Yes
Observations	6,691	6,691	6,691
R^2	0.010	0.019	0.025
Mean of Dep. Variable	0.489	0.489	0.489

Notes: Attendance rate is defined as actual attendance divided by the stadium's maximum capacity. Standard errors clustered at the team-pair level (home $\times$ away) are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

References

Abramitzky, Ran, Liran Einav, Shimon Kolkowitz, and Roy Mill, "On the Optimality of Line Call Challenges in Professional Tennis," *International Economic Review*, 2012, *53* (3), 939–964.

Anbarci, Nejat, Jungmin Lee, and Aydogan Ulker, "Win at All Costs or Lose Gracefully in High-Stakes Competition? Gender Differences in Professional Tennis," *Journal of Sports Economics*, 2016, *17* (4), 323–353.

Dufner, Anna-Lena, Lisa-Marie Schütz, and Yannick Hill, "The Introduction of the Video Assistant Referee Supports the Fairness of the Game – An Analysis of the Home Advantage in the German Bundesliga," *Psychology of Sport and Exercise*, 2023, *66*, 102386.

Holder, Ulrike, Thomas Ehrmann, and Arne König, "Monitoring Experts: Insights from the Introduction of Video Assistant Referee (VAR) in Elite Football," *Journal of Business Economics*, 2022, *92* (2), 285–308.

Hsu, Mike, "Umpire Home Bias in Major League Baseball," *Journal of Sports Economics*, 2024, *25* (4), 423–442.

Mills, Brian and Rodney Fort, "League-Level Attendance and Outcome Uncertainty in U.S. Pro Sports Leagues," *Economic Inquiry*, 2014, *52* (1), 205–218.

Mills, Brian M., "Policy Changes in Major League Baseball: Improved Agent Behavior and Ancillary Productivity Outcomes," *Economic Inquiry*, 2017, *55* (2), 1104–1118.

Parsons, Christopher A., Johan Sulaeman, Michael Yates, and Daniel S. Hamermesh, "Strike Three: Discrimination, Incentives, and Evaluation," *American Economic Review*, 2011, *101* (4), 1410–1435.

Paulsen, R. J., "Temporary employment and the protection of investments in human capital: examining the Major League Baseball player market," *Economic Inquiry*, 2025, pp. 1–12.

Pope, Devin G., Joseph Price, and Justin Wolfers, "Awareness Reduces Racial Bias," *Management Science*, 2018, *64* (11), 4988–4995.

Price, Joseph and Justin Wolfers, "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

Smith, Erin E. and Jon D. Groetzinger, "Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games," *Journal of Quantitative Analysis in Sports*, 2010, *6* (1), 4.

Thrane, Christer, "Nationalistic Bias in Experts' Player Ratings in Football," *Economics Letters*, 2025, *247*, 112187.