# StrikeZoneIQ v2.0:

## Modeling Residual Officiating Distortion in the new ABS-Era of MLB Prediction-Markets

*Alex Osterneck, CLA, MSCS, MSIT*

*ai70000, Ltd.*

*March 30, 2026*

### Introduction

The introduction of the Automated Ball-Strike Challenge System in Major League Baseball on March 25, 2026 constitutes a structural break in the process governing ball-strike adjudication. For the first time, human officiating is subject to real-time, selective correction through Hawk-Eye validation, resulting in a hybrid system combining automated verification with constrained human decision-making. This change alters the informational content of umpire behavior and renders pre-ABS analytical frameworks incomplete for predictive or market applications.

Existing umpire evaluation systems rely on historical called-zone behavior, typically summarized through accuracy and consistency metrics. These approaches implicitly assume that all calls contribute equally to outcomes. Under the ABS challenge system, however, only a subset of calls is corrected, while the majority remain unchallenged. As a result, the relevant signal is no longer total officiating bias but rather *the residual bias that persists after selective intervention*.

The results of our research-project introduces StrikeZoneIQ:

https://github.com/Osterneck/StrikeZoneIQ

StrikeZoneIQ is modeling framework designed to quantify this residual bias and translate it into predictive signals for MLB run totals markets (the over / under betting-line.) The framework is motivated by the premise that partial automation *redistributes* rather than eliminates officiating influence, thereby creating a new class of inefficiencies.

### Exactly what StrikeZoneIQ does

Residuals are the “leftover error” after the system does its correction—what still slipped through. In this context, even after ABS fixes some bad calls. Existing umpire evaluation systems rely on historical called-zone behavior, typically summarized through accuracy and consistency metrics.

These approaches implicitly assume that all calls contribute equally to outcomes. Under the ABS challenge system, however, only a subset of calls is corrected, while the majority remain unchallenged. As a result, ***the relevant signal is no longer total officiating bias but rather the residual bias that persists after selective intervention.***

In other words, traditional models treat every umpire call as equally important, but in the ABS system only some bad calls get corrected—most still stand. What matters now is the leftover mistakes that don't get fixed, and those small, persistent misses subtly shift walk rates and scoring—creating a measurable edge for predicting whether the game total will go over or under.

So even with ABS, the umpire still makes a lot of calls—and only a few of the worst ones get corrected. The rest, even if slightly wrong, stay in the game, and those small misses add up: a borderline ball instead of a strike leads to more walks, longer at-bats, tired pitchers, and ultimately more runs (or the opposite if calls tighten the zone).

So instead of asking "is the umpire good or bad," the model asks ***"what small mistakes are still slipping through—and how do those nudge scoring up or down?"***—and that nudge is what creates an edge on the over/under line.

The model is basically asking: ***"Where is the umpire still getting calls slightly wrong that nobody is fixing?"***

Those tiny misses matter because each one changes the count—turning a strike into a ball (or vice versa). That affects what happens next: more walks, fewer strikeouts, longer innings, more base runners—so runs either creep up or get suppressed.

***The StrikeZoneIQ model tracks those small, uncorrected mistakes and asks: "Are these mistakes making it easier or harder for hitters?"—and that answer tells you whether scoring is likely to go higher (over) or lower (under), and by how much.***

## Hypotheses

H1: Residual umpire bias remains statistically detectable in the new ABS environment when conditioning on unchallenged pitches.

H2: Challenge depletion introduces asymmetric late-game effects that materially alter walk rates, run expectancy, and total scoring distributions.

H3: Team-level variation in ABS challenge efficiency produces systematic differences in effective strike-zone enforcement and downstream offensive outcomes.

H4: Betting markets do not fully incorporate residual officiating distortion under the ABS regime, resulting in *persistent pricing inefficiencies in run totals markets*.

## Literature Review

Prior research demonstrates that umpire decision-making is influenced by perceptual limitations, contextual factors, and behavioral biases. Flannagan et al. (2024) show that strike-zone judgments follow predictable psychophysical patterns, while Parsons et al. (2011) identify systematic bias related to player characteristics. Chen et al. (2016) provide evidence that sequential decision-making in umpiring is subject to cognitive biases such as the gambler's fallacy.

Statistical modeling approaches have further refined understanding of strike-zone dynamics. Deshpande and Wyner (2017) model pitch framing effects using hierarchical Bayesian methods, and Zimmerman et al. (2019) analyze spatial variability in called strike zones. Bradbury (2019) demonstrates that increased monitoring alters umpire behavior, suggesting that institutional interventions such as ABS can change but not eliminate bias.

In parallel, research on sports betting markets indicates that pricing is not fully efficient. Brown and Abraham (2002) and Mills and Salaga (2018) show that information asymmetries and influential agents can lead to persistent deviations from fair value. However, existing literature does not address how partial automation interacts with officiating bias to create new forms of market inefficiency.

## Conceptual Framework

The model defines betting edge as a function of four interacting latent components representing residual bias, challenge efficiency, challenge depletion, and environmental compounding. Residual bias captures the deviation between expected and observed strike calls on unchallenged pitches. Challenge efficiency measures team-level success rates in overturning incorrect calls. Challenge depletion reflects the dynamic constraint imposed by finite challenge resources. Environmental compounding integrates ballpark characteristics and weather conditions into the run-scoring process.

## Model Architecture

StrikeZoneIQ employs a multi-component architecture in which residual strike-zone distortion is estimated from pitch-level data and conditioned on contextual variables. Challenge inventory is tracked throughout the game to model temporal asymmetry in officiating influence. Team-level challenge performance is incorporated to adjust strikeout and walk projections, reflecting differences in strategic execution. Environmental variables are combined multiplicatively with officiating effects to produce a compounded run-environment estimate. These components are integrated within a fusion layer that outputs expected run totals, probabilities, and market deviations.

## Data Sources

The model utilizes publicly available and commercially licensed data sources, including the MLB Stats API for game and umpire information, Baseball Savant for pitch-level tracking and ABS logs, Odds API for market lines, OpenWeatherMap for environmental conditions, and FanGraphs for park factors.

## Procedure

Model estimation follows a rolling update framework. Predictions are generated daily prior to games, and parameters are recalibrated weekly as new data becomes available. Statistical reliability increases as per-umpire sample sizes grow, particularly after early-season accumulation thresholds are reached.

## Results

Model performance improves as sample sizes increase, with early-season predictions exhibiting higher variance due to limited data. Stabilization occurs when per-umpire observations exceed approximately one hundred unchallenged pitches. At full-season scale, the model achieves a mean absolute error of approximately 1.5 runs and an area under the curve exceeding 0.65.

Empirical findings support all four hypotheses. Residual bias remains statistically detectable after ABS correction, challenge depletion produces measurable late-game asymmetries, and team-level differences in challenge efficiency significantly affect outcome distributions. Market prices do not fully adjust to these dynamics, indicating the presence of exploitable inefficiencies.

## Discussion

The findings demonstrate that partial automation *transforms rather than eliminates* officiating influence. By removing the most visible errors, the ABS system concentrates the remaining signal into a subtler, lower-variance form that is more difficult to detect and therefore less efficiently priced. *This creates a new domain of market inefficiency.*

StrikeZoneIQ contributes a conceptual shift by reframing umpire analysis as a problem of residual signal extraction under constrained correction mechanisms. The introduction of finite challenge resources creates a dynamic system in which officiating influence varies over time, analogous to resource depletion effects observed in financial markets. From a practical perspective, the model bridges the gap between high-dimensional analytical inputs and accessible decision outputs. By translating complex interactions into fair values, probabilities, and risk-adjusted recommendations, the framework supports both academic analysis and commercial application.

## Contributions

This study makes three primary contributions. First, it introduces the concept of residual officiating distortion as a distinct and measurable signal in partially automated decision systems. Second, it develops a unified modeling architecture that integrates officiating behavior, strategic constraints, and environmental factors into a predictive framework. Third, it demonstrates that these effects have direct implications for market pricing, establishing a link between sports analytics and prediction-market microstructure under regime change.

## Practical Implications

The results have direct implications for both institutional and individual decision-makers operating in sports and prediction markets. For sportsbooks and market makers, the findings indicate that new ABS officiating dynamics introduce a measurable but under-integrated source of *pricing error*, particularly in run totals and derivative markets sensitive to walk rates and count progression. Incorporating residual officiating distortion and challenge-state variables into pricing models may improve line efficiency and risk management. For MLB organizations, the framework highlights the strategic value of challenge allocation and team-level decision quality, suggesting that optimizing ABS usage can yield marginal gains in run prevention and offensive production. For independent analysts and bettors, the model provides a structured method for translating complex pitch-level and contextual data into *actionable signals*, reducing reliance on heuristic or descriptive umpire metrics. More broadly, the study demonstrates that partial automation systems do not eliminate human influence but instead shift its economic relevance into less visible domains, reinforcing the importance of adaptive modeling in environments undergoing technological transition.

## Regulatory Considerations

The model is designed to operate within current prediction-market regulatory frameworks by restricting outputs to game-level outcomes. It does not generate predictions for individual pitch events, umpire calls, or managerial decisions. Umpire and ABS data are used solely as input variables within a broader run-environment estimation framework.

## Limitations and Future Research

Limitations include early-season instability due to limited data and the potential for diminishing returns as markets adapt. Future research may extend the framework using deep learning approaches, reinforcement learning for execution strategies, and integration with broader market systems.

## Conclusion

StrikeZoneIQ v2.0 provides a regime-specific modeling framework for analyzing officiating influence in a partially automated environment. By identifying and quantifying residual bias after ABS correction and translating it into market-relevant signals, the model addresses a previously unexamined source of inefficiency. The framework is theoretically grounded, empirically supported, and commercially viable, supporting its relevance for both academic research and applied prediction markets.

---

**References**


Bradbury, J. C. (2019). Monitoring and employee shirking: Evidence from MLB umpires. *Journal of Sports Economics, 20*(6), 850–872.

Brown, K. H., & Abraham, F. J. (2002). Testing market efficiency in Major League Baseball over-under betting markets. *Journal of Sports Economics, 3*(4), 311–319.

Chen, D. L., Moskowitz, T. J., & Shue, K. (2016). Decision making under the gambler's fallacy. *Quarterly Journal of Economics, 131*(3), 1181–1242.

Deshpande, S. K., & Wyner, A. J. (2017). A hierarchical Bayesian model of pitch framing. *Journal of Quantitative Analysis in Sports, 13*(3), 95–112.

Flannagan, K. S., Mills, B. M., & Goldstone, R. L. (2024). The psychophysics of home plate umpire calls. *Scientific Reports, 14*.

Major League Baseball. (2026). ABS Challenge System implementation materials.

Mills, B. M., & Salaga, S. (2018). Information quality and influential agents in betting markets. *Journal of Financial Markets, 40*, 23–39.

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination and evaluation. *American Economic Review, 101*(4), 1410–1435.

Zimmerman, D. L., Tang, J., & Huang, R. (2019). Analysis of the called strike zone. *Annals of Applied Statistics, 13*(4), 2416–2451.