THE UNIVERSITY OF CHICAGO

Does crisis turn into opportunity? The effect of manager challenge in Major League Baseball

By

Hyun Ki Kim

June 2021

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Devin Pope
Preceptor: Min Sok Lee

Does crisis turn into opportunity?

The effect of manager challenge in Major League Baseball

Hyun Ki Kim

The University of Chicago

May 27, 2021

Abstract

This study estimates the effect of manager challenge in Major League Baseball to test whether momentum exists in sports. Despite much research over the last 30 years, the existence of momentum in sports remains controversial. Productivity of teams, measured by the number of runs in the inning, is analyzed by using 1,193 replay reviews during 2015-2019 regular season American League home games. Using a difference-in-differences methodology, this study fails to find statistically significant causal effect of winning a manager challenge on the team's productivity on subsequent innings. However, there could still be a shorter- or longer-term momentum as this study analyzes the existence of a medium-term momentum that lasts for subsequent two innings. The finding could be generalized to other zero-sum settings where agents have option to challenge.

1. Introduction

Momentum is originally a term in physics, but it is commonly used in finance and sports to describe the tendency of good or bad outcomes happening subsequently. However, determining whether momentum exists is a challenge due to the following reasons. The underlying motive for momentum is related to psychological factors such as self-efficacy and motivation. Also, many factors that contributed to earlier outcomes affect later outcomes as well. Therefore, a randomized experiment is needed to determine the existence of a momentum.

Sports is generally a good setting to test human behaviors (Palacios-Huerta 2014). Performance of experienced agents that have high stakes are measured with detail and they have well prespecified goals and rules. Therefore, sports can be viewed as a large-scale field experiment, and the evidence from sports can be generalized to other human behaviors.

Researchers have tried to find evidence of momentum, but no clear evidence has yet been found. Thus, momentum has been, in general, regarded as a fallacy by scholars. However, recent study (Miller and Sanjurjo 2018) suggests that studies on the earlier studies have long been miscalculating the causal effect of momentum. This scholarly debate on momentum calls for research that relies on rigorous causal analysis. Against this backdrop, this study, using a unique natural experiment, empirically tests whether momentum exists in sports.

The modern technology of the video assistant referee (VAR) system serves as an ideal natural experiment. In a perfectly rational world, the result of a VAR reviews should not affect the player's behavior, as it was the right call that should have been made. However, the challenge winning team would consider it as a gain while the opponent as a loss because close calls are seen as having been able to go the other way. Therefore, a VAR reviews can serve as a critical point that initiates a momentum.

This study will analyze whether the productivity of a team changes when the team wins a VAR review. Baseball is an ideal setting to test productivity for structural and empirical reasons. Structurally, baseball has a clear separation between the offense and the defense. Therefore, it is possible to measure the productivity of each team separately, unlike in other sports. As for the empirical reasons, huge amounts of high-quality baseball data are publicly available, ensuring a high-quality analysis.

Two opposing hypotheses as to the effects of winning VAR reviews have been suggested by scholarship. A positive momentum would increase productivity for the challenge-winning team while a negative momentum would decrease productivity for the challenge-losing team. Using the unique baseball data, this paper empirically tests these hypotheses.

2. Background

After Gilovich et al. (1985) found no evidence of the "hot hand" in basketball, researchers believed that the momentum in sports was merely a misperception of randomness. However, Miller and Sanjurjo (2018) argued that these studies that treated momentum as misperception were subject to streak selection bias and that the "hot hand" in fact exists. Therefore, a reexamination is needed to test whether the momentum effect in sports exists.

Baseball is played between two teams with 9 alternating offense opportunities for each team that are called innings. In each inning, the goal of offense team is to score as many runs as possible before they accumulate three outs that brings the inning to an end. The result of a baseball game is determined by the cumulative runs between two teams.

There is a home team and an away team in a baseball game, and the away team always starts the offense first. Innings where away team plays offense is called the top of the inning, and

where home team plays offense is called the bottom of the inning. If the home team is winning after the top of the ninth inning, the game is finished without playing the bottom of the ninth inning. However, if the cumulative runs are same for both teams after the ninth inning, extra innings are played until the game winner is determined.

With the technological improvement, instant replay review system was implemented in Major League Baseball (MLB) in 2014 and has been modified in 2015 (MLB replay review regulations). Managers may request a video replay in certain cases, after an umpire has made a call, and this request is called a challenge. Managers are limited to one challenge per game, but if a request they made is successful—that is the original call was incorrect—the manager retains the right to make another challenge. Umpires also have an option to initiate a replay review to check rules and to confirm home runs without restriction but may initiate replay review for other plays only after the 6^{th} inning.

In order to challenge a call, managers must notify the umpire within 30 seconds after the play. Each team has technical staff who has access to replays that help managers decide whether to challenge or not. If the calls are subject to replay review, then the game is paused, and two umpires put headphones on to communicate with Replay Command Center in New York City.

The result of a replay review is determined solely by the Replay Command Center, which has access to tens of cameras from each stadium. The center aims to decide within two minutes, and the call from the field stands if there is no clear and convincing evidence from the video replay. Therefore, there are three outcomes for a replay reviews that is overturned, confirmed, and stood. However, players and managers cannot distinguish between the calls that were confirmed and those that were stood, because they can only see the result of a play being overturned or upheld.

After the adoption of the current MLB replay review system in 2014, two studies have examined the system. Imber (2015) found a positive correlation between challenge-winning and game-winning. Furthermore, the study found that challenges made on weekends were more likely to be overturned. Marx Scheuerell and Marx (2019) examined the factors effecting the percentage of calls overturned. They found that challenges made on earlier innings and bases were more likely to be overturned. Although prior work on replay reviews suggest potential for a momentum effect, further study is needed to move forward from correlation to causation.

3. Data

Play-by-play data on 2015-2019 regular season American League home games was collected from Baseball-Reference.com. In addition, supplementary information on the replay reviews were collected from the Retrosheet website (retrosheet.org). MLB consists of two leagues, but in order to minimize the interrelationship between offense and defense, this study will focus on the American League home games where pitchers do not participate during the offense. Prior to the 2015 season, managers had a maximum of 2 challenges per game and thus their future usage was restricted even if both challenges were successful. Because of this deprivation in option value, data were collected starting from the 2015 season. Also, postseason games were excluded because managers start games with 2 challenges and the sample size is relatively small.

The dataset consists of 108,537 innings and 522,229 plays. Out of 3,235 replay reviews, 2,853 reviews (88%) were initiated by managers, and 382 reviews (12%) by umpires. After an average of 90 seconds of review, 1,566 calls (48%) were overturned, and 1,669 calls (52%) were upheld. However, the sample of this study will be restricted to the 1,193 replay reviews that were

initiated by managers between top of the 3rd inning and bottom of 6th inning to compare the productivity of innings before and after a manager challenge.

Baseball-Reference.com provides detailed information on any given situation during a match. Those include variables such as inning, run, batting order, home team, away team, date, game result, pitcher, batter, runner, score, error, out count, pitch count, runner left on base, win expectancy, win probability added, pitcher change, pinch hit, pinch run, batting outcome, as well as information on replay reviews.

Since the managers of both the offense and the defense team can challenge, being overturned or upheld itself does not imply being favorable to one team. Therefore, offense challenged-overturned and defense challenged-upheld are combined as offense challenge win (defense challenge lose), offense challenged-upheld and defense challenged-overturned are combined as offense challenge lose (defense challenge win). Out of 16 innings that had 2 challenges in the sample, 12 innings that had same offense challenge win remained in the sample and 4 innings that had different offense challenge win were removed.

Table 1 summarizes the descriptive statistics of the key variables in the sample. Vast majority of the sample had a single challenge, with an average duration of 95 seconds. 45% of the challenges turned out to be favorable to offense team. Innings were restricted to 3rd to 6th inning with about 51% played in the bottom of the inning. The score difference, which was calculated by subtracting the defense team score from the offense team score, had mean of -0.14. There was on average 1 runner left on base at the end of the inning, 0.13 home run, and 0.69 strike out per inning in the sample.

Table 1: Summary statistics of innings in the sample

	Observations	Mean	Std. dev.	Min	Max
Challenge	1193	1.01	0.10	1	2
Challenge duration (in seconds)	1181	95.31	44.60	21	361
Offense challenge win	1193	0.45	0.50	0	1
Inning	1193	4.54	1.12	3	6
Home game	1193	0.51	0.50	0	1
Score difference (offense - defense)	1193	-0.14	2.70	-13	12
Runner left on base	1193	1.01	0.86	0	3
Error	1193	0.15	0.40	0	2
Home run	1193	0.13	0.38	0	3
Strike out	1193	0.69	0.75	0	3

4. Methodology

This study aims to test whether winning a manager challenge has a causal effect on the productivity of subsequent innings. Although it is possible to measure the productivity of a team in various ways, runs in inning is chosen as a dependent variable for the following reasons. First, the result of a baseball game is determined by the aggregate runs over 9 innings. The objective of a baseball team during a game would be scoring more runs than their opponent, and therefore runs in inning would be a better measure than, for example, hits in inning or on-base percentage. Another alternative would be to use batter-level productivity to test the existence of a short-term momentum. However, as the existence of the next batting opportunity depends on the result of a manger challenge—it is more likely to have subsequent batting opportunity when the offense won

the challenge. Due to this selection bias, batter-level productivity is not suitable for difference-in-differences methodology.

To compare two innings before and after a manager challenge, sample was restricted to challenges that happened during 3rd to 6th inning. Difference-in-differences methodology was used to compare runs in innings before and after a manager challenge.

The model used in analyzing the effect for offense team is the following:

$$\begin{aligned} Inning\ run_i = {} & \beta_0 + \beta_1\ Offense\ challenge\ win_i + \beta_2\ D_{t-2,i} + \beta_3\ D_{t,i} + \beta_4\ D_{t+1,i} + \beta_5\ D_{t+2,i} \\ & + \beta_6\ Offense\ challenge\ win_i \cdot D_{t-2,i} + \beta_7\ Offense\ challenge\ win_i \cdot D_{t,i} \\ & + \beta_8\ Offense\ challenge\ win_i \cdot D_{t+1,i} + \beta_9\ Offense\ challenge\ win_i \cdot D_{t+2,i} \\ & + \sum_{j=10}^{38} \beta_j\ Team_{j,i} + \varepsilon_i \end{aligned}$$

where i indicates individual inning, $Inning\ run_i$ is number of runs scored, $Offense\ challenge\ win_i$ is a dummy variable indicating whether the challenge that occurred in the game that the $Inning\ run_i$ is in was favorable to offense, $D_{T,i}$ is a dummy variable indicating relative distance from the challenge, $Team_{j,i}$ is a dummy variable for each team in MLB, and ε_i is a random error. Dummy variable $D_{t-1,i}$ and $Challenge\ win_i \cdot D_{t-1,i}$ were omitted as a base case. Therefore, it is possible to test whether there are any differences in productivity between $t-1$ and $t-2, t, t+1, t+2$ when the manager challenges were favorable and against offense.

The model used in analyzing the effect for defense team is the following:

$$\begin{aligned} Inning\ run_i = {} & \beta_0 + \beta_1\ Defense\ challenge\ win_i + \beta_2\ D_{t-1.5,i} + \beta_3\ D_{t+0.5,i} + \beta_4\ D_{t+1.5,i} \\ & + \beta_5\ Defense\ challenge\ win_i \cdot D_{t-1.5,i} + \beta_6\ Defense\ challenge\ win_i \cdot D_{t+0.5,i} \\ & + \beta_7\ Defense\ challenge\ win_i \cdot D_{t+1.5,i} + \sum_{j=8}^{36} \beta_j\ Team_{j,i} + \varepsilon_i \end{aligned}$$

where i indicates individual inning, $Inning\ run_i$ is number of runs scored, $Offense\ challenge\ win_i$ is a dummy variable indicating whether the challenge that occurred in

the game that the $Inning\ run_i$ is in was favorable to defense, $D_{T,i}$ is a dummy variable indicating relative distance from the challenge, $Team_{j,i}$ is a dummy variable for each team in MLB, and ε_i is a random error. Dummy variable $D_{t-0.5,i}$ and $Challenge\ win_i \cdot D_{t-0.5,i}$ were omitted as a base case. Therefore, it is possible to test whether there are any differences in productivity between $t - 0.5$ and $t - 1.5, t + 0.5, t + 1.5$ when the manager challenges were favorable and against defense.

This identification strategy assumes that the effect of winning a challenge to last for two subsequent innings. It is also assumed that the challenge winning and challenge losing situations have parallel trend of expected runs for two innings before and after the challenge.

5. Results

The result of the estimation for the offense team is illustrated in Figure 1 and Table 2. The estimation for defense team in Figure 2 and Table 3. Heteroskedasticity robust standard errors were used and control variables such as batting order, score difference, and home game status were included to test the sensitivity of the analysis.

In terms of offense team, the only difference between the challenge-winning condition and the challenge-losing condition occurred at the t inning. That is β_7 in the model for offense was estimated to be meaningfully different from zero. However, in terms of defense team, there were no meaningful difference between the challenge-winning condition and the challenge-losing condition.

Figure 1: Runs in inning by offense challenge win

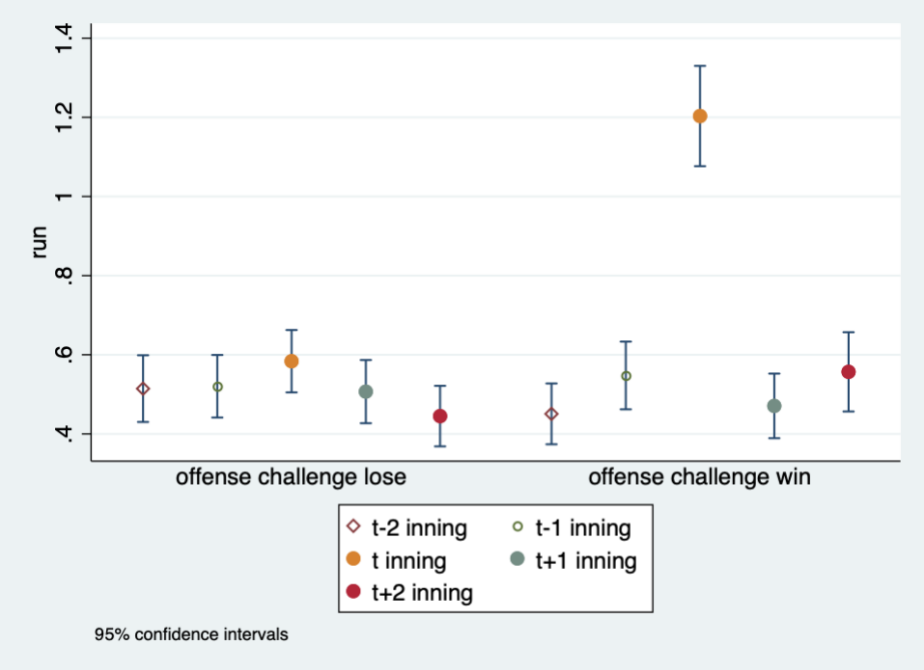


Figure 2: Runs in inning by defense challenge win

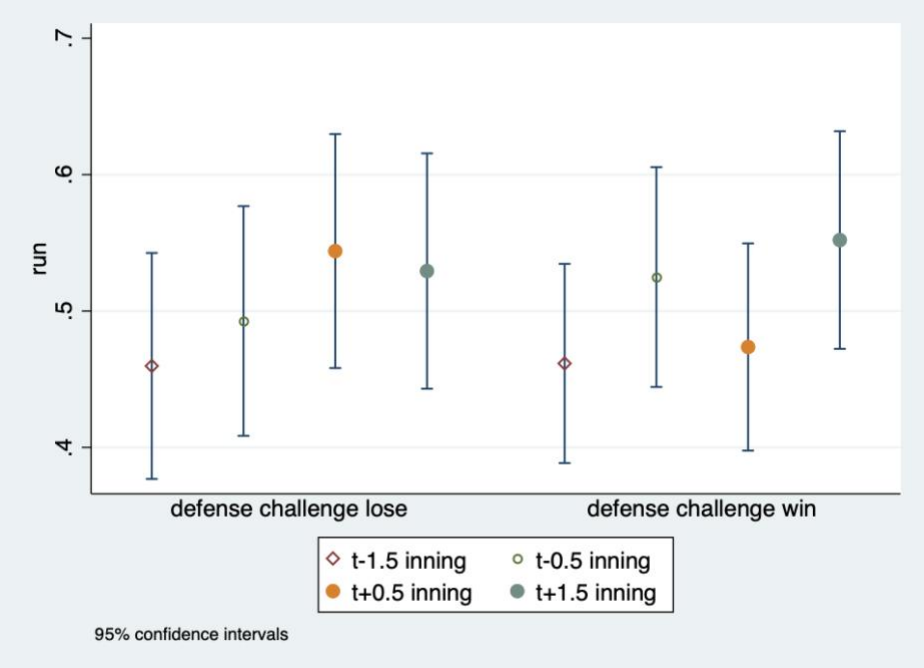


Table 2: Effect of offense challenge win on runs in inning

	(1) run	(2) run	(3) run	(4) run
Offense challenge win	0.0358	0.0339	0.0339	0.0343
	(0.62)	(0.59)	(0.59)	(0.60)
D_{t-2}	0.0154	-0.00297	-0.00304	-0.00303
	(0.29)	(-0.06)	(-0.06)	(-0.06)
D_t	0.0702	0.0706	0.0705	0.0706
	(1.29)	(1.30)	(1.30)	(1.30)
D_{t+1}	0.0320	0.0282	0.0284	0.0284
	(0.60)	(0.53)	(0.53)	(0.53)
D_{t+2}	-0.0413	-0.0432	-0.0430	-0.0430
	(-0.80)	(-0.84)	(-0.83)	(-0.83)
Offense challenge win $\cdot D_{t-2}$	-0.120	-0.118	-0.118	-0.118
	(-1.51)	(-1.48)	(-1.48)	(-1.48)
Offense challenge win $\cdot D_t$	0.583***	0.584***	0.585***	0.584***
	(6.07)	(6.10)	(6.10)	(6.10)
Offense challenge win $\cdot D_{t+1}$	-0.114	-0.112	-0.112	-0.112
	(-1.41)	(-1.38)	(-1.38)	(-1.38)
Offense challenge win $\cdot D_{t+2}$	0.0213	0.0247	0.0252	0.0251
	(0.25)	(0.29)	(0.30)	(0.30)
Batting order start		-0.0168***	-0.0168***	-0.0168***
		(-3.30)	(-3.30)	(-3.31)
Score difference			-0.000914	-0.000847
			(-0.18)	(-0.16)
Home game				0.0202
				(0.72)
_cons	0.0751	0.174	0.172	0.172
	(0.58)	(1.36)	(1.35)	(1.35)
N	5965	5965	5965	5965

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Heteroskedasticity robust standard errors were used

Team fixed effects are omitted to conserve space

Table 3: Effect of defense challenge win on runs in inning

	(1) run	(2) run	(3) run	(4) run
Defense challenge win	-0.0256	-0.0257	-0.0260	-0.0243
	(-0.44)	(-0.44)	(-0.45)	(-0.42)
$D_{t-1.5}$	-0.0711	-0.0729	-0.0734	-0.0732
	(-1.26)	(-1.29)	(-1.30)	(-1.30)
$D_{t+0.5}$	0.0317	0.0311	0.0346	0.0349
	(0.55)	(0.54)	(0.60)	(0.60)
$D_{t+1.5}$	0.0320	0.0314	0.0347	0.0350
	(0.55)	(0.54)	(0.59)	(0.60)
Defense challenge win $\cdot D_{t-1.5}$	0.0276	0.0269	0.0270	0.0272
	(0.35)	(0.34)	(0.34)	(0.35)
Defense challenge win $\cdot D_{t+0.5}$	-0.0729	-0.0729	-0.0761	-0.0761
	(-0.91)	(-0.91)	(-0.95)	(-0.95)
Defense challenge win $\cdot D_{t+1.5}$	0.0130	0.0127	0.00984	0.00991
	(0.16)	(0.16)	(0.12)	(0.12)
Batting order start		-0.00391	-0.00403	-0.00358
		(-0.72)	(-0.74)	(-0.66)
Score difference			0.00450	0.00466
			(0.84)	(0.87)
Home game				0.0798**
				(2.73)
_cons	1.072***	1.095***	1.090***	1.086***
	(3.69)	(3.74)	(3.73)	(3.72)
N	4772	4772	4772	4772

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Heteroskedasticity robust standard errors were used

Team fixed effects are omitted to conserve space

6. Discussion

The result of the analysis suggests that there does not exist medium-term momentum that is initiated by the manager challenge. For the offense, inning where offense won the challenge had 0.58 higher runs on average, but the difference in differences between challenge winning and losing for $t-1$ and $t-2, t+1, t+2$ were not meaningfully different from zero. For the defense, the difference in differences between challenging winning and losing for $t-0.5$ and $t-1.5, t+0.5, t+1.5$ were all not meaningfully different from zero. This implies winning a manager challenge does not have meaningful causal effect that last for subsequent two innings for both offense and defense.

The estimates are stable for adding control variables such as batting order, score difference, and home game status. However, only the starting batting order of the inning for the offense is meaningful in predicting the number of runs per inning among control variables. Since different teams have different abilities to score a run, the estimation includes team-fixed effects. Furthermore, models were estimated under heteroskedasticity assumption.

The effects are stable across using different measures for the productivity. Appendix Figures 1 – 4 used hits in inning, on-base percentage, and win probability added as a dependent variable. The results are also stable across using different sample innings. Appendix Figure 5, 6 used challenges in 2nd to 7th inning and 4th to 5th inning respectively.

One might be concerned that the null result from the estimation could be originating from a low statistical power. If the standard errors were large, it would not be possible to find the effect even if it exists. For offense, if there were momentum effect that lasts for $t+1, t+2$ innings, the model would be able to find any effect larger than approximately 30% of run differences in the t inning.

Although the study did not find any evidence of momentum effect, the result could be dependent on sample selection. The sample were chosen to be the nearby innings in 2015-2019 regular season American League home games that have challenges in 3rd to 6th inning. Therefore, the causal effect estimated by the model would be momentum in the middle inning of regular season American League games. One could still argue momentum may exist in post season games, National Leagues games, innings at the beginning or at the end of the game. Also, shorter- or longer-term momentum such as momentum for subsequent batters or subsequent games may still exist even if there were no momentum for subsequent innings.

Also, the parallel trend assumption for the difference-in-differences may be violated if the original call by the umpires is biased. Studies found that sports referees are subject to many different types of biases (Chen et al. 2006; Garicano et al. 2005; Moskowitz and Wertheim 2011; Price and Wolfers 2010). For example, the parallel trend assumption may be violated if umpires had incentive to minimize the score difference to keep the game more engaging for the fans.

7. Conclusion

The existence of momentum in sports has been highly controversial among researchers, players, managers, and sports fans for over 30 years. This study used manager challenge as a natural experiment to test this controversial hypothesis. But the findings fail to find momentum effects for subsequent innings. Hopefully, this study can be a cornerstone for analyzing momentum in sports and in other zero-sum games that have components like a managerial challenge in baseball. For example, studying the effect of a challenge on productivity in sports team may be generalized to lawsuits between duopoly companies.

References

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. "Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires." *The Quarterly Journal of Economics* 131, no. 3 (2016): 1181-1242.

Garicano, Luis, Ignacio Palacios-Huerta, and Canice Prendergast. "Favoritism under social pressure." *Review of Economics and Statistics* 87, no. 2 (2005): 208-216.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. "The hot hand in basketball: On the misperception of random sequences." *Cognitive Psychology* 17, no. 3 (1985): 295-314.

Imber, Gil. "Reviewing instant replay: Observations and implications from replay's inaugural season". *Baseball Research Journal* 44, no. 1 (2015): 45-53.

Marx Scheuerell, Anne C., and David B. Marx. "An examination of MLB play call challenges." *Baseball Research Journal* 48, no. 1 (2019): 49-52.

Miller, Joshua B., and Adam Sanjurjo. "Surprised by the hot hand fallacy? A truth in the law of small numbers." *Econometrica* 86, no. 6 (2018): 2019-2047.

MLB replay review regulations. *Major League Baseball*. Accessed May 27, 2021. http://m.mlb.com/glossary/rules/replay-review.

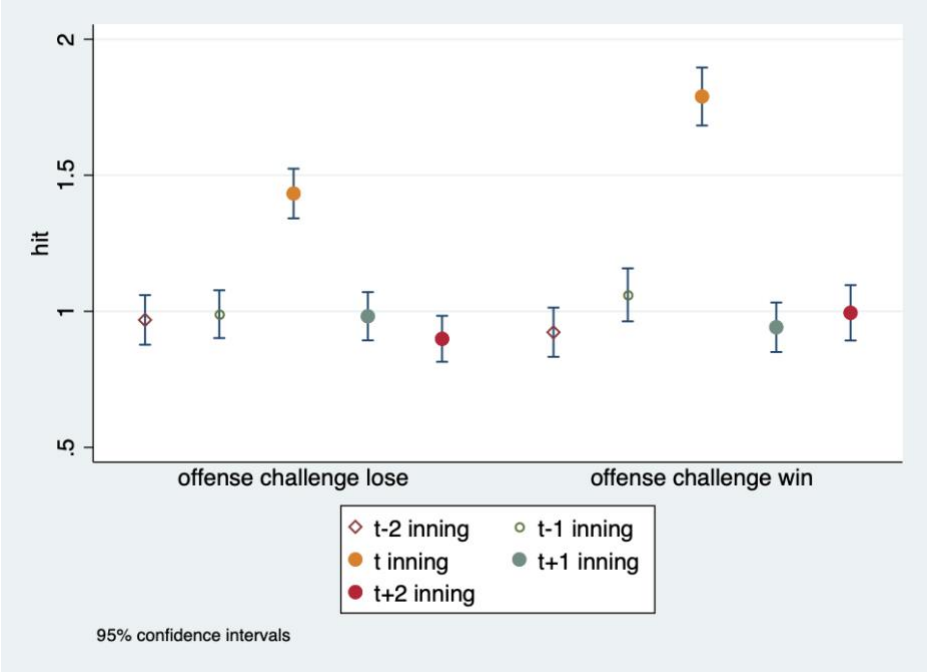
Moskowitz, Tobias J., and L. Jon Wertheim. *Scorecasting: The hidden influences behind how sports are played and games are won*. Crown Archetype, 2011.

Palacios-Huerta, Ignacio. *Beautiful game theory*. Princeton University Press, 2014.

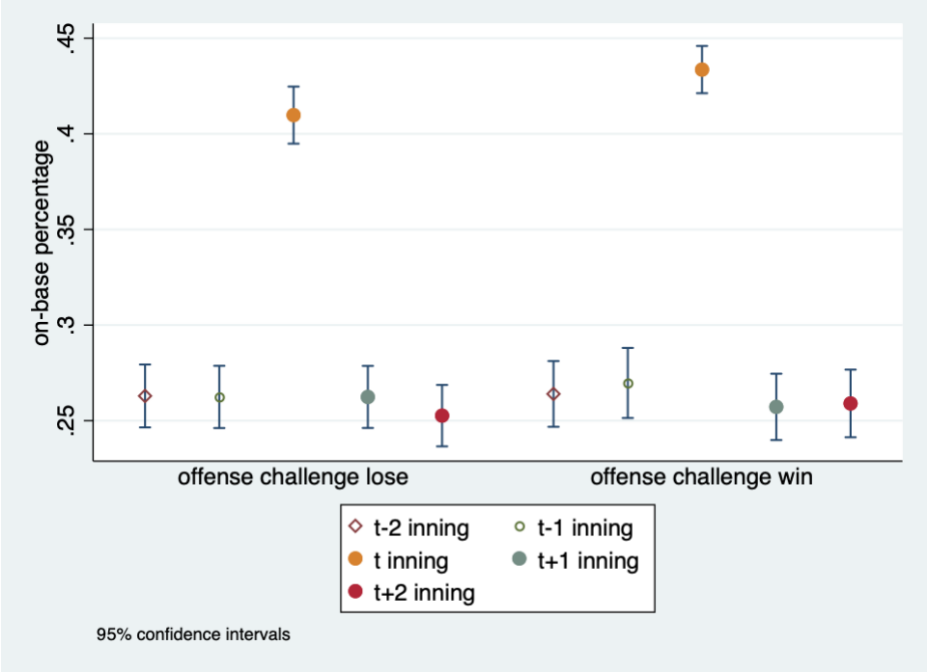
Price, Joseph, and Justin Wolfers. "Racial discrimination among NBA referees." *The Quarterly Journal of Economics* 125, no. 4 (2010): 1859-1887.

Appendix

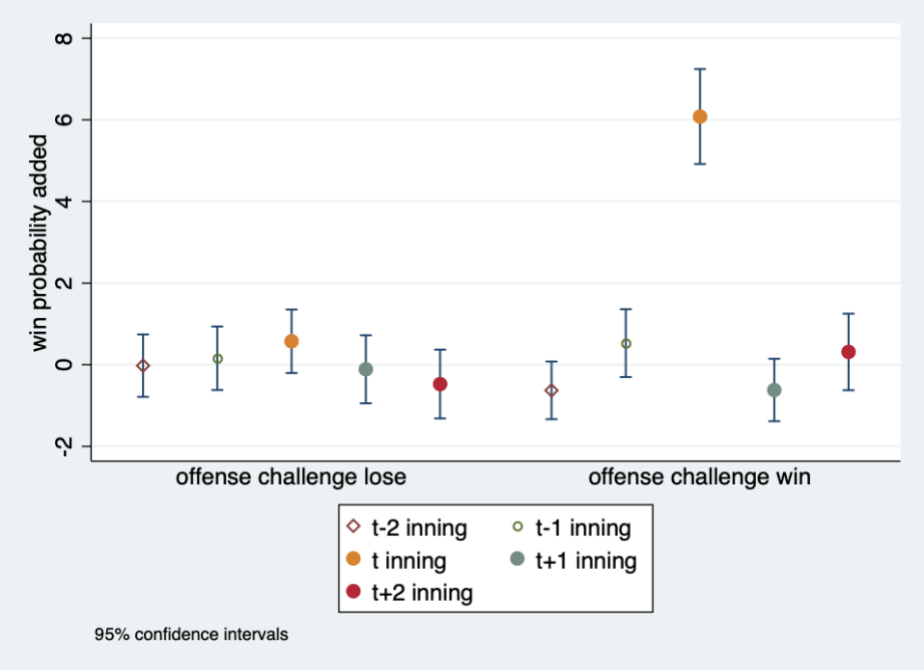
Appendix Figure 1: Hits in inning by offense challenge win



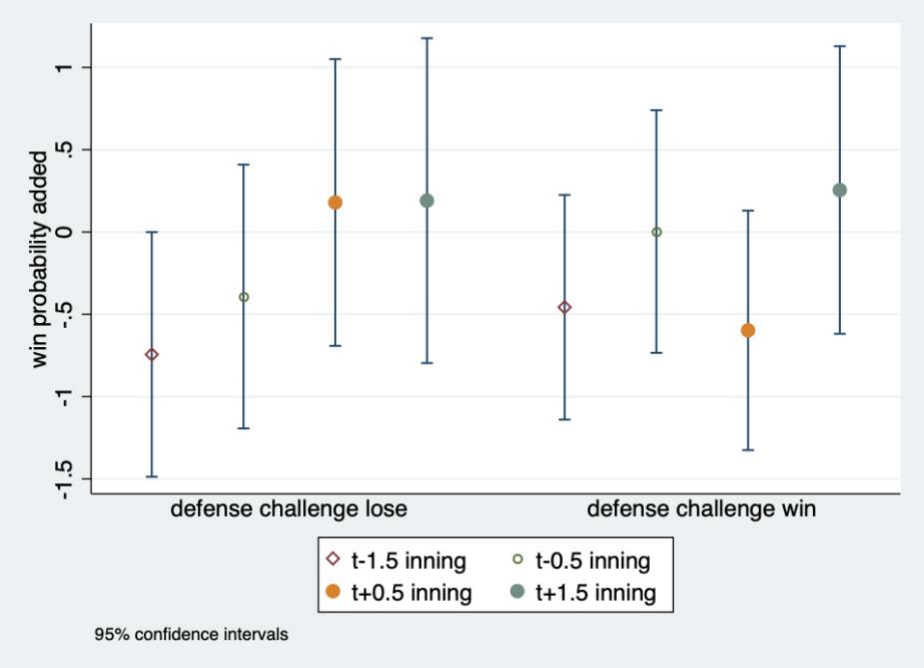
Appendix Figure 2: On-base percentage in inning by offense challenge win



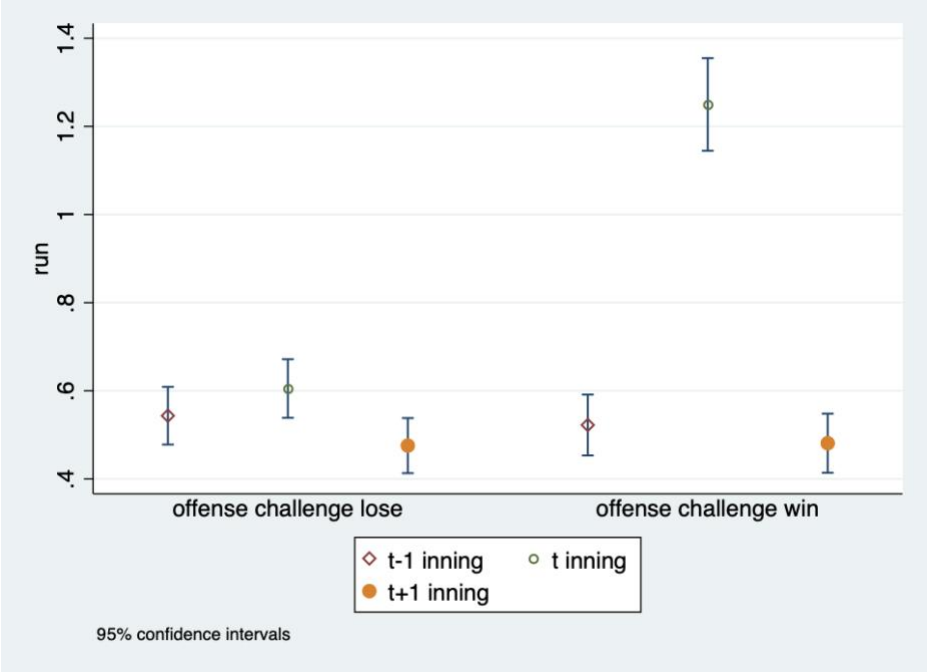
Appendix Figure 3: Win probability added in inning by offense challenge win



Appendix Figure 4: Win probability added in inning by defense challenge win



Appendix Figure 5: Runs in inning by offense challenge win (2nd to 7th inning)



Appendix Figure 6: Runs in inning by offense challenge win (4th to 5th inning)