Senior Project

Department of Economics

Field of Dreams: Will They Come for MLB's Quicker, More Thrilling Games?

Carson Keenan

Spring 2024

Advisor: *Dr. Ali Enami*

Abstract

In 2023, Major League Baseball introduced rule changes with the goals of decreasing game length and increasing game excitement. This paper analyzes the impact these changes had on attendance, as well as the impact of similar changes at the minor league level eight years prior. The MLB rule changes included a pitch clock, a ban on certain defensive shifts, and an increase in base size, all of which can increase game pace and excitement and theoretically increase attendance numbers. Using a one-way fixed effects log-linear model, this analysis finds that the 2023 rule changes increased average Major League attendance by 8.4% at 10% significance level. Conversely, a two-way fixed effects difference-in-differences log-linear model shows that the earlier rule changes at the minor league level had no statistically significant impact on attendance. This discrepancy suggests that the quality and visibility of MLB games amplify the impact of these rule changes. This paper recommends that Major League Baseball continues to test new rules at the minor league level, since currently these tests seem to have no significant economic impact. Future research on the Major League side should include more post-rule change seasons to confirm these results, especially in the post-pandemic era.

Table of Contents

I. Introduction

In 2023, Major League Baseball introduced several changes to the rules of the game with the goals of increasing excitement and decreasing game duration. One rule change gives specific guidelines as to where fielders can align themselves on defense. This effectively bans the use of the "shift," a defensive deployment in which an extra infielder moves to the opposite side of the field. This change ideally benefits hitters, as it reduces methods in which the defense can use analytics against them. Another new rule increases the size of the bases, which is thought to incentivize stealing bases. It also serves as a form of player safety, as the larger area provides more space for fielders and baserunners to share the bases without sustaining injuries from collisions.

The most significant adjustments, however, target the pacing of the game itself. These rules establish a pitch clock, limiting the amount of time allowed between pitches and at bats. In addition to this requirement, pitchers are limited to two disengagements from the pitching rubber per at bat. Further disengagements would result in a balk, allowing baserunners to advance one base. Batters are required to be ready in the batter's box by the time the pitch clock reads 9 seconds; otherwise, they receive an automatic strike. Furthermore, batters may only call timeout once per at bat.

While this was the first season with these rule changes at the Major League level, they had been tested at the Minor League level several years prior. In 2015, the pitch clock debuted in the minors. The goal was to shorten the average game length, and this method was successful at first. Eventually, average game length began to increase again. While pitch clocks did limit the amount of time per pitch, pitchers could artificially extend this time indefinitely by disengaging from the pitching rubber. There was no limit on how many times a pitcher could do this at the time, and each disengagement reset the pitch clock. To correct this, restrictions on pickoff throws and batter timeouts were introduced in 2022. These corrections had a notable impact, as average game time saw significant decreases across all levels of minor league baseball.

Why does Major League Baseball want to speed up the game? Commissioner Rob Manfred believes that longer game times bore fans, especially within the younger generation. In fact, all the new rules lend themselves to more exciting baseball in theory: more steals, more hits, and faster games. In theory, higher attendance numbers also benefit both the local and national economies. Increased attendance implies more ticket sales, but also leads to increases in concessions and memorabilia sales, benefiting the teams. The government likewise benefits from these sales, as they can collect taxes on these transactions. Restaurants and stores in the immediate area could experience more activity, as fans would likely visit before or after the game. On a more regional level, businesses such as sports bars may also notice an increase in business, as increased fandom may drive people to go out and watch televised games even when not attending in person. The question, however, is whether these changes matter to fans. In other words, have the rule changes in Major League Baseball influenced game attendance numbers?

While previous papers in this area individually study determinants of attendance and impact of the 2023 MLB rule changes on player statistics, they have yet to measure the direct result of the rule changes on MLB attendance. This analysis aims to find the impact of the MLB rule changes on attendance, as well as determine whether or not similar rule changes affect minor league attendance.

This paper employs four datasets. The first contains the MLB specific data relevant to this analysis, including attendance figures, team winning percentages, payroll, and average game length. The second contains similar data for each minor league team at the AAA level. The third and fourth contain the population for the nearest Metropolitan Statistical Area to each team in both leagues: American cities and Canadian cities require the use of two separate sources for population data.

This paper analyzes the impact of similar rule changes at both the Major League and minor league levels. For the Major Leagues, this analysis uses a one-way fixed effects log-linear model to measure the impact of the 2023 rule changes on attendance. It contains multiple versions of the model, with varying amounts of controls. This paper also contains a two-way fixed effects difference-in-differences log-linear model to examine the effect similar rule changes have on minor league attendance. Since a relaxed version of the pitch clock was introduced in 2015, this model allows for more years of post-treatment analysis. Furthermore, Major League data serves as a control in this model. In 2022, a stricter pitch clock was introduced; however, and so acts as a second treatment. This paper finds a parallel trend between the data, asserting that Major League and minor league attendance follow a similar trend before the treatment period. All regressors have insignificant differences between both groups aside from population; this difference cannot be avoided, as Major League teams play in significantly larger cities than minor league teams on average.

The Major League analysis finds that the 2023 rule changes have a positive impact on attendance, while the minor league analysis shows no significance. In the best Major League model, the rule changes introduce an 8.4% increase in average attendance across Major League Baseball. This number remains consistent even after removing the regressor for population, an insignificant control variable. The minor league models suggest that this impact does not translate between leagues, as both versions of the pitch clock have negligible influence on average attendance.

Why did the results vary drastically between the Major League and minor league models? Since Major League Baseball is a higher quality product, it follows that ticket prices for these games are more expensive on average. Fans will likely be more considerate about a more expensive product, allowing improvements to the quality of the aforementioned product to have a larger impact. Furthermore, the MLB uses the minor leagues as a test subject for their rule changes. If they improve the on-field statistics that they believe correlate with higher attendance, the changes are then implemented at the Major League level.

The positive impact on attendance measured in the Major League model should not be interpreted without caution. Firstly, this model lacks year fixed effects, as having only one year of post-treatment data makes this type of model impossible. While a control for the first post-COVID-19 season (2021) is included in the model, this may not fully account for the rebound in attendance seen as a result of society moving back towards normal functionality. Furthermore, one year of post-treatment data is capable of being a statistical anomaly, and the result is only significant at 10% significance level. Another analysis of this nature in the future with more data after the rule changes will provide a more accurate representation of the long term effects, as future data would be less affected by both pandemic related behaviors and any potential bias towards the recency of the treatment.

Despite these limitations, the MLB model is in line with the expected results, while the minor league model shows no significant impact. With these results, the MLB should continue analyzing what statistical patterns drive attendance. They should also utilize the minor leagues as an experiment for potential rule changes. If these changes positively impact statistics and receive positive reviews from players, then it is worth considering implementation at the highest level.

The rest of this paper is organized in the following manner: Section II discusses determinants of attendance and statistical impacts of the rule changes; Section III rules describes the data used in this analysis; Section IV explains the theory behind changing the rules; Section V defines the methodology; Section VI provides the results, and Section VII concludes.

II. Literature Review

Many researchers theorize on the determinants of sports attendance worldwide, and this is especially prevalent in soccer, the world's most popular sport. In their research on Maracanã, a Brazilian soccer stadium, Barajas and Gasparetto (2023) find that ticket price has a significant negative impact on attendance, which is in line with the theory of supply and demand. This effect, however, is dependent on which sector of the stadium is being analyzed. Unlike American stadiums, where fans of either team can typically purchase tickets for any seat in a facility, this Brazilian stadium (like European stadiums) has designated sectors for home and away fans. In addition to price, the analysis shows that the home team's probability of winning has a significant impact on attendance in certain sectors. Home fans are more likely to attend games in which their team is favored. While the same is true of away fans, they are also more likely to purchase tickets as huge underdogs.

Closer to home, Major League soccer features a slightly different stadium seating structure, and also varies in its attendance determinants. Bradbury (2020) finds that Major League Soccer attendance is largely driven by performance and novelty. Attendance has a positive relationship with goal differential, although this effect experiences diminishing returns. Success in the previous season also positively correlates with attendance. As for novelty, the two significant factors to attendance are team age and soccer-specific stadiums. Newer teams bring in more fans than older teams, while soccer-specific stadiums see more fans than multi-purpose stadiums. Market, which is measured by population, income, and the presence of marquee players, is not a significant contributor to MLS attendance.

While some determinants of attendance are shared across sports, unique factors to baseball attendance numbers certainly exist. In their analysis on minor league baseball attendance, Gitter and Rhoads (2010) find that winning percentage has a significant impact on attendance at the A and AA levels. Minor league attendance was also positively associated with the ticket price of local Major League teams, representing a substitution effect. Interestingly, minor league attendance had a positive relationship with the winning percentage of a local Major League only when the minor league team was an affiliate of the Major League team. At the Major League level, attendance in the modern era leans toward exciting gameplay. While winning percentage is certainly a significant factor, attendance numbers are also positively correlated with game uncertainty and league-wide offensive performance (Ahn and Lee, 2014). In essence, fans are more likely to attend a close game with many extra base hits and home runs.

At the Major League level, specific team factors play a role in attendance numbers as well. Lim and Pedersen (2022) find that the away team's payroll and history of championships have a positive affect on attendance. This result is indicative of the realization of two potential theories. First of all, teams that are frequently successful and have more popular, superstar players (represented by a higher payroll) incentivize fans to travel to away games, and may be more likely to have fans in cities other than their home. In addition, home fans invest more in attending games against stronger opponents in the hope that their team can pull off an important victory. This also mimics the results found at the minor league level, where attendance positively correlate with uncertainty of outcome (Ahn and Lee, 2014).

Game-specific variables also play a role in determining Major League attendance. The time of games significantly contributes to attendance, as more fans go to night games than day games (Denaux et al., 2011). In addition, they find that games played on the weekend (Friday-Sunday) draw the largest attendance numbers, with Saturday games being the most popular. Night games do not have to compete with people working standard hour occupations like day games do, while work and school are less likely to interfere with weekend games in general. Another game specific factor is the pitching matchup. Ormiston (2014) observes that star pitchers on both home and away teams have a positive impact on attendance. While the focus is on offense in more recent years, superstar players at any position encourage attendance, whether they are on the home team or not. Overall; however, home teams benefit on the field from increased attendance numbers, as they gain a statistically significant advantage to their win percentage (Smith and Groetzinger, 2010).

Rule changes to increase excitement and decrease game duration have already been tested at the Minor League level, and the results on player statistics are significant. Starting in 2021, High-A (a lower-level Minor League division) required left-handed pitchers to disengage from the pitching rubber before throwing to first base for a pickoff attempt. Prior to this, it was much more difficult for baserunners to discern whether a left-handed pitcher was delivering a pitch or attempting to pick them off. In turn, this made stealing second base off a left-handed pitcher more challenging prior to the rule change. High-A baseball also limited the number of pickoff attempts per at bat to 3, one of the rule changes that was introduced at the Major League level last season. How did these changes impact steals? Between 2019 and 2021, the success rate on steals at the High-A level increased from 66.43% to 75.81%, which is significant. Furthermore, the mean value of stolen base attempts per game rose significantly from 1.22 to 1.59 in the same time span (Houghtaling 2022).

Due to the staggered release of the rule changes between Minor League Baseball and Major League Baseball, some players gained experience playing with the new rules sooner than others. Hitters that had previous experience with the new rules demonstrated a significant advantage over inexperienced hitters across multiple offensive categories, while pitchers in the same circumstances exhibited negligible improvements (Shaw 2024). This statistical difference between hitters and pitchers indicates that the rule changes may be more taxing on pitchers than hitters.

III. Preview of the Data

This paper utilizes data from four main sources. Specific Major League baseball data, including attendance numbers, team payrolls, and winning percentages, is collected from Baseball Reference (Sports Reference LLC., 2024). Minor league baseball data is collected from The Baseball Cube (The Baseball Cube, 2024). United States population data is collected from the U.S. Census Bureau (U.S. Census Bureau, 2024). Lastly, population data for the Toronto metropolitan area is collected from Statistics Canada (2024).

Baseball Reference is a database dedicated to collecting and tracking a variety of statistics, facts, and figures on baseball at multiple levels, especially Major League Baseball. The database contains annual data for multiple variables relevant to this paper. These variables include each Major League team's average attendance per game, payroll, average game length, winning percentage, and playoff appearances. This paper uses 18 years of data for each of these variables, from 2005 to 2023. A line graph of Major League Baseball's annual average attendance is visible in Figure 1 as part of a parallel trend test, located in Section V of this paper. From this graph, it is apparent that from 2007 to 2019 (before COVID-19), MLB attendance follows a decreasing trend. Note that no data point exists for 2020 and that 2021 is a low outlier. Due to the COVID-19 pandemic, baseball games had zero attendance in 2020 and limited attendance in 2021 (dependent on individual stadium capacity regulations).

The Baseball Cube is another large database containing a variety of numbers regarding baseball. This paper specifically uses attendance data from minor league teams at the AAA level, the highest level of Minor League Baseball. This data spans 17 seasons from 2005 to 2022, and in addition to attendance data, contains winning percentage and playoff appearances. As with the previous database, there is no data from 2020 due to the pandemic.

Population data for Metropolitan Statistical Areas closest to each Major League team is provided by two databases. This paper utilizes data from the U.S. Census Bureau for American metropolitan population. For the population of the Toronto and Ottawa Metropolitan Statistical Areas, data from Statistics Canada is used. Each of these sources provide estimates from July of each calendar year. This paper uses 11 years of census data for most cities, from 2012 to 2023. A select number of cities did not yet have population estimates for 2023; this paper estimates these values with the 2021-2022 population trend.

Table 1 contains the summary statistics for all quantitative variables utilized in the Major League model, as well as the average game length. The standard deviations of payroll and population are quite large when compared to their respective means. Since Major League Baseball contains two teams in New York City, Chicago, and Los Angeles, their population data points appear twice for every year in this data set. Furthermore, payroll in Major League Baseball is uncapped, so large outliers in spending have no true maximum. Another intriguing data point in this set is the minimum game length value of 2.633 (2 hours, 38 minutes). This data point specifically comes from 2023 when the rule changes were in play. This value is roughly 28 minutes shorter than the average game length from 2012-2022, which is nearly a 15% decrease. This demonstrates that the rule changes induce a drastic reduction in game length.

Table 1. Major League Summary Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Average Attendance	28,486.13	8,744.16	7,934	49,066
Win Percentage	0.500	0.078	0.290	0.685
Team Payroll	$124,605,134	$52,739,239	$14,672,300	$265,140,429
Average Game Length (hours)	3.063	0.146	2.633	3.417
Population (MSA)	6,142,860.16	4,591,553.77	1,559,792	19,774,386

Sources: Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics Canada (2024).

Table 2 summarizes the data for all minor league data that the second model utilizes. Compared to Major League teams, average attendance is significantly lower. This results from several factors, including lower capacity stadiums and less player/brand recognition. In American sports, minor leagues or developmental leagues in general are seen in a lower regard than the top leagues, much different than the pyramid league structure popular in Europe, for example. Teams cannot promote and relegate between leagues in the United States (a minor league team cannot become a Major League team). The standard deviation of win percentage is lower at the minor league level, indicating that the skill difference between the best and worst AAA teams is less pronounced than at the Major League level. Minor league teams also play in lower population areas on average. The large standard deviation hints at some exceptions that play in the same Metropolitan Statistical Areas as Major League teams, such as Tacoma (Seattle's AAA affiliate) and Sugar Land (Houston's AAA affiliate).

Table 2. Minor League Summary Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Average Attendance	6,298.07	1,739.57	1,714	10,497
Win Percentage	0.500	0.059	0.340	0.662
Population (MSA)	1,607,940.72	1,211,734.35	419,261	7,947,439

Sources: The Baseball Cube (2024), U.S. Census Bureau (2024), Statistics Canada (2024).

IV. Theoretical Framework

Past data gives reason to believe that faster, more exciting games may increase attendance by decreasing the opportunity cost of attending a game. Over the past 10 years, game length trended upwards. MLB games last considerably longer than NBA games, and until last season rivaled NFL games in length. When games last longer, the opportunity cost of attending a game also increases. As a result, Major League attendance followed a negative trend in the same time span. Compared to football and basketball, baseball has noticeably more "dead time", or time in which no action is actively happening on the field. Longer games with no increase in entertainment value cause some fans to replace in-person attendance with their next best alternative.

To combat this, the rule changes in 2023 aim to eliminate the "dead time" between pitches, at bats, and innings. Prior to 2023, "dead time" would drastically increase during crucial, game-deciding batter-pitcher matchups. These at bats frequently contained numerous utilizations of step offs by the pitcher and timeouts called by the batter, as both players sought to gain an edge. While this can build intensity in playoff situations, it can become excessive during regular season play. At its peak, minutes can pass without a pitch actively being thrown. For the 2023 season, Major League Baseball attacks both of these issues directly. Firstly, a pitch clock was implemented, which affects both pitchers and batters. Pitchers are required to start their motion prior to the clock hitting zero; otherwise, a ball is awarded to the batter. The batter is required to be in the batter's box and looking at the pitcher by the time the clock hits nine seconds; else, the pitcher receives a free strike. On top of this, the MLB greatly reduces the ability to prolong these matchups with tactical interruptions. Batters are limited to one timeout per at bat, while pitchers are allotted two disengagements. (A disengagement is any time that a pitcher steps off or "disengages" from the pitching rubber, usually in an attempt to pick off a baserunner.) Theoretically, these changes to the pacing of at bats will allow for fewer stoppages and more engaging game action.

In addition to speeding up the game, Major League Baseball implements adjustments designed to increase the likelihood of some of baseball's most exciting plays: steals. For the 2023 season, base sizes increased from 15 square inches to 18 square inches. The MLB claims that the purpose of this change is namely a safety concern, and this may be true. It does; however, bring advantages to baserunners attempting to steal. The increase in size means that bases are slightly closer together. While the change may seem minimal, the difference between being out or safe on a stolen base attempt can be a game of inches. Not only does the size increase allow baserunners to arrive to the base slightly sooner, it also grants them more flexibility to avoid a tag by the fielder while still making contact with the base. As stated earlier, these changes have already made an impact in minor league testing: both steal attempts and successes are on the rise (Houghtaling 2022). Coupled with the decreased time between pitches, fans could be treated to an increase in thrilling steal attempts with even less time between each one.

Finally, the MLB aims to bolster both individual offensive and defensive efforts through the banning of the shift. In recent years, a defensive tactic referred to as “the shift” has become a dominating strategy, especially against power hitters. In general, right-handed power hitters tend to pull the ball to the left half of the field, while left-handed hitters do the opposite. This tendency also applies when they do not hit the ball squarely, resulting in a ground ball instead of a line drive. To counter this, defenses began shifting to have three infielders on that side. This meant that power hitters had to choose between sacrificing power and attempting to hit the ball to the opposite side, or swing for the fences with a much greater likelihood of making an out if they missed. Last season, Major League Baseball effectively banned this strategy by requiring two infielders on either side of second base at all times. This allows power hitters to play to their strengths, as they can swing for power more often while still being able to get hits through holes in the defense that the shift can no longer cover. This does make defense more difficult, but it does allow the MLB’s exceptional athletes more chances to make exquisite individual plays in the field. This change, along with larger bases, provides baseball with the fuel necessary to make the game more entertaining. With the pitch clock increasing the pace of play as well, shorter games with more frequent exciting plays ideally decrease the opportunity cost of going to a game and incentivize more fans to attend.

V. Methodology

This analysis contains two different models to evaluate the impact of rule changes on in person attendance. The first model strictly looks at the impact at the Major League level, and measures how much attendance changes due to the 2023 rule changes. The second model measures these effects at the minor league level, utilizing the Major League teams as a control group.

The Major League model utilizes a one-way fixed effect log-linear model, and the equation is as follows:

$$Log(Attendance_s) + B_0 + B_1 RuleChange + \ X_s + Team_s + Year + Year^2 + Year^3 + \varepsilon_s \ \ (1)$$

Log(Attendance) measures the average per game attendance for team *s*, and is taken in logarithmic form. *RuleChange* is an indicator variable equal to 1 in year 2023, when the MLB rule changes took effect. *X* represents several control variables, including winning percentage, previous season's playoff appearance, team payroll, whether the team played in a new ballpark in a given year, and population within the nearest metropolitan statistical area. One control also indicates the year 2021, the season immediately following the height of the COVID-19 pandemic. All quantitative control variables are taken in a log form in this model. *Team* represents team fixed effects. *Year,* along with its associated polynomial terms, represent the reference year. The reference year is defined as the number of years since the height of COVID-19 in 2020. In this definition, years before COVID-19 are negative values, while years after COVID-19 are positive values. Lastly, ε is the white noise.

The minor league analysis utilizes a two-way fixed effect difference-in-differences model, and the equation is as follows:

$$Log\left(Attendance_{s,t}\right) = B_0 + B_1 RuleChange1_{s,t} + B_2 RuleChange2_{s,t} + X_{s,t} + Team_s + Year_t + \varepsilon_{s,t} \ (2)$$

Log(Attendance) measures the average per game attendance for team *s* in year *t*, and is taken in logarithmic form. *RuleChange1* is an indicator variable that represents the first level of rule changes applied for each team. *RuleChange2* is an indicator variable that represents the second level of rule changes applied for each team. For all Major League Teams, this value is 0 in all years, since they are the control group. (This model only goes through 2022, so it does not include the year the MLB rule changes took effect.) For Minor League teams, the coding is as follows. Prior to 2015, both treatment variables are 0. From 2015 to 2021, *RuleChange1* is 1, while *RuleChange2* remains at 0. In 2022, *RuleChange1* changes back to 0, and *RuleChange2* changes to 1. These values represent the different levels of Minor League rule changes: a more relaxed pitch clock was used from 2015 to 2021, and was replaced with a stricter pitch clock and additional rule changes in 2022. In essence, the minor league teams experience two treatments. *X* represents several control variables, including winning percentage, previous season's playoff appearance, and population within the nearest metropolitan statistical area. *Team* and *Year* represent team and year fixed effects, respectively. Lastly, ε is the white noise.

To test the similarity of the control variables between the minor league and Major League data, this analysis conducts a balance of regressors test. The results of this test appear in Table 3. The controls of win percentage and making the playoffs last year are nearly identical, and their differences are statistically insignificant. Population; however, is significantly larger in the control group on average at 1% significance level. This result is expected, as Major League teams tend to play in significantly larger cities than their minor league counterparts. Overall, Major League and minor league teams are similar entities aside from the locations in which they play.

Table 3. Balance of Regressors Test

Regressors	Treatment	Control	Difference
Win Percentage	0.5000	0.4974	0.0026
Metropolitan Statistical Area Population	1387419	5865113	-4477694***
Playoff Last Year	0.2634	0.2800	-0.0166

Source: Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics Canada (2024) and own calculations.
Notes: The table shows the pre-2015 average values of regressors for the treatment and control groups as well as their differences. Differences that are statistically significant are identified with *, **, and *** corresponding to 10%, 5%, and 1% significance levels, respectively.

To test the similarity in behavior between Major League and minor league attendance, this paper utilizes a parallel trend test. A visual representation of this test is shown in Figure 1. Minor league attendance is significantly lower on average than Major League attendance, and the changes from year to year are more difficult to see at the minor league level. It seems that multiple behaviors could be occurring here. Prior to 2015, it appears that the minor league attendance loosely follows the trend of Major League attendance. At some points; however, the data appears to possibly be moving in opposite directions, signaling a potential substitution effect. To further examine this, the analysis includes a statistical parallel trend test. The results are displayed in Table 4. From the statistical test, it is clear that minor league and Major League follow a parallel trend, as the interaction terms between the year polynomial and treatment have no statistical significance.

Figure 1. Parallel Trend Visual Test

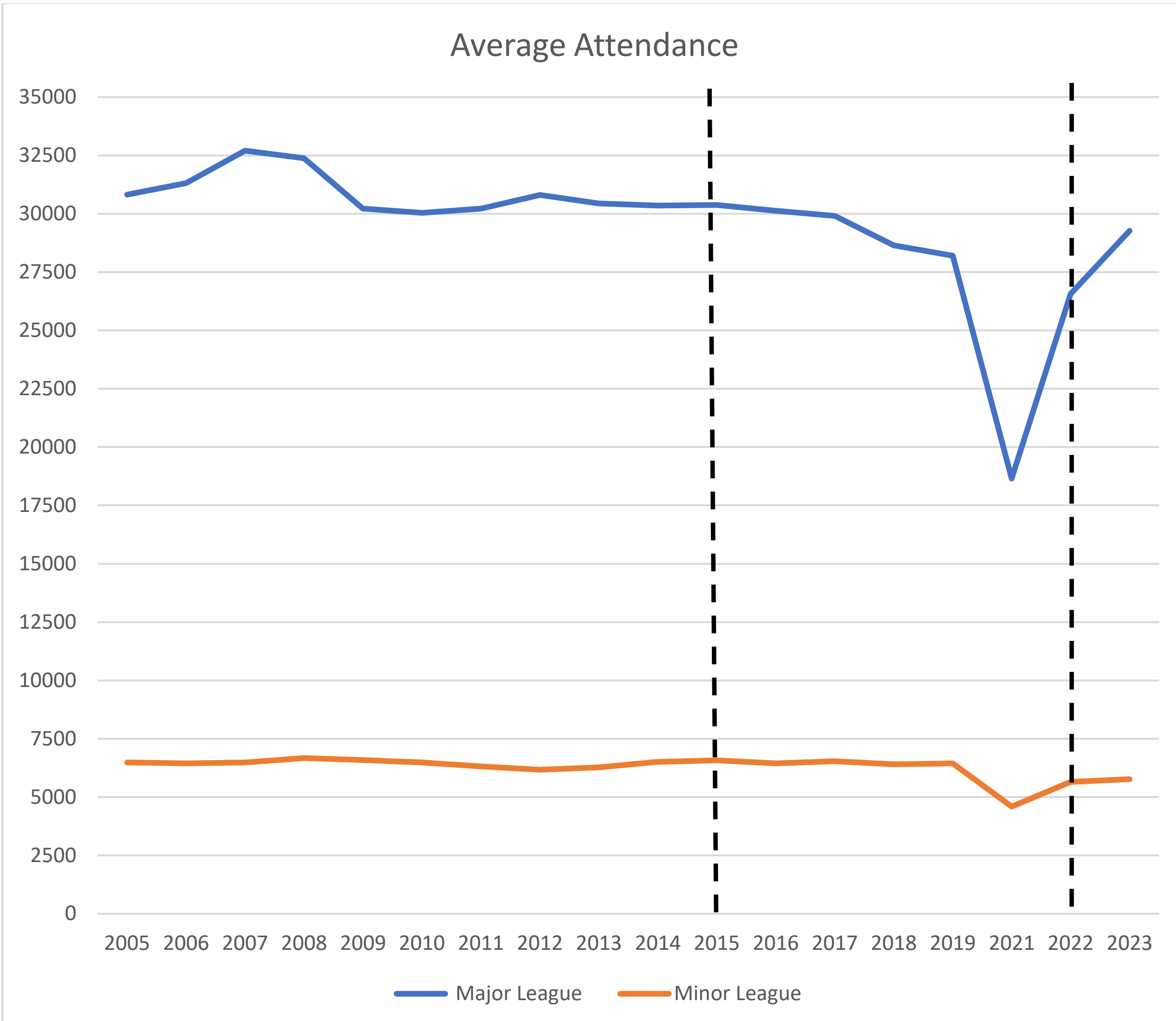

Table 4. Parallel Trend Statistical Test

Variables	Model1	Model2
Treatment	-1.88	-1.43
	(1.43)	(1.36)
Log(Metropolitan Statistical Area Population)		0.10***
		(0.02)
Log(Win Percentage)		0.53***
		(0.10)
Previous Season Playoff Appearance		0.12***
		(0.03)
Relative Year	0.24	0.21
	(0.28)	(0.24)
$(\text{Relative Year})^2$	0.02	0.02
	(0.03)	(0.02)
$(\text{Relative Year})^3$	0.00	0.00
	(0.00)	(0.00)
Treatment*Relative Year	-0.10	-0.00
	(0.45)	(0.42)
Treatment* $(\text{Relative Year})^2$	-0.01	-0.00
	(0.04)	(0.04)
Treatment* $(\text{Relative Year})^3$	-0.00	0.00
	(0.00)	(0.00)
Intercept	11.04***	9.70***
	(0.91)	(0.84)
Number of Observations	543	543
Adjusted R-Square	0.8801	0.9019
Overall Significance	600.29***	537.06***

Source: Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics Canada (2024).
Notes: Robust Standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels respectively.

VI. Results

The Major League analysis finds that in general, the 2023 rules changes had a statistically significant positive effect on attendance. The complete results for each model can be found in Table 5. The analysis is broken into three models: one with no control variables (except for the COVID-19 indicator variable), one with all control variables, and one with every control variable except for population (which was removed due to statistical insignificance). This paper refers to

Model 3 as the best model. In this model, the rule changes increase average Major League attendance by 8.4%, on average, at 10% significance level. This is economically significant as well, as this converts to an average increase of about 2,400 ticket sales per game, or 194,400 ticket sales per season. This result is in line with the MLB's vision of bringing more fans back to baseball stadiums. Their hope is that shorter, more exciting baseball games would not only appeal to longtime baseball fans, but to newer or more casual fans of the game as well (who may have been less likely to attend games in their previous state). These early results suggest that the theory behind these changes is correct, although a long-term analysis is certainly necessary. In the first season after COVID-19 (2021), the best model finds that average attendance is, on average, 38% less than what would have been expected. This is due to most stadiums not operating at full capacity for most or all that season. Teams playing in a new ballpark experienced a 36% increase in average attendance their first year inside that ballpark, on average. This is a small sample size; however, as only 3 teams moved to new ballparks during the sample period. Teams who made the postseason in the previous season received, on average, a 9% boost to average attendance compared to if they had missed the playoffs. A 1% increase in payroll leads to a 0.18% increase in average attendance, while a 1% increase in win percentage corresponds to a 0.37% increase in average attendance, on average. All control variables referenced above are significant at 1% significance level.

Table 5. Major League Model Results

Variables	Model1	Model2	Model3
Rule Changes	0.136*	0.084*	0.084*
	(0.073)	(0.045)	(0.045)
Post-Covid Year (2021)	-0.37***	-0.38***	-0.38***
	(0.05)	(0.04)	(0.04)
Log(Metropolitan Statistical Area Population)		-0.17	
		(0.29)	
Log(Payroll)		0.19***	0.18***
		(0.03)	(0.03)
Log(Win Percentage)		0.38***	0.37***
		(0.05)	(0.05)
New Ballpark		0.36***	0.36***
		(0.13)	(0.13)
Previous Season Playoff Appearance		0.09***	0.09***
		(0.02)	(0.02)
Relative Year	-0.03***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)
$(\text{Relative Year})^2$	0.0007	0.0024	0.0025
	(0.0041)	(0.0026)	(0.0026)
$(\text{Relative Year})^3$	0.0003	0.0003	0.0003
	(0.0004)	(0.0003)	(0.0003)
Intercept	10.04***	9.55**	6.92***
	(0.03)	(4.43)	(0.53)
Number of Observations	330	330	330
Adjusted R-Square	0.7480	0.8755	0.8758
Overall Significance	63.04***	95.64***	95.01***

Source: Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics Canada (2024).
Notes: Robust Standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels respectively. This model contains Team fixed effects.

The minor league analysis finds that neither the relaxed pitch clock nor the strict pitch clock are statistically significant determinants of attendance. This analysis contains two models: one without controls, and one with controls. Table 6 summarizes the results of these models. The model with controls is the better model. In this model, the relaxed pitch clock leads to a 1.8% increase in attendance, while the strict pitch clock actually results in a 4.2% decrease in attendance. As stated before; however, neither of these coefficients are statistically significant, so the effect of the rule changes in minor league baseball is effectively zero. Similar to the Major League model, a postseason appearance in the previous season leads to a 7% increase in average attendance, on average. A 1% increase in population correlates with a 0.96% increase in average attendance, while a 1% increase in win percentage results in a 0.32% increase in average attendance, on average. These control variables are all significant at 1% significance level. The outcome of statistical insignificance for rule changes at the minor league level is not in line with the hopes Major League Baseball had for attendance numbers. This discrepancy in the minor league and Major League models is expected; however, as the rule changes had different goals at each level. Rule changes at the minor league level are typically in a testing phase, and Major League Baseball uses them to gather data on the statistical and analytical effectiveness of the rule changes on the gameplay itself. If the rule changes have the desired effect, they are then introduced at the Major League level. Since the gameplay statistics are proven, the Major League level then becomes a better measure on whether or not these changes to the gameplay actually impact attendance. Fans do not tend to have affiliations with minor league teams nearly as strongly as they do Major League teams. More often than not, minor league attendance is a result of baseball fans having a team in the area (potentially significantly closer than any Major League team). They may be more likely to attend if the team is an affiliate of their closest Major League team as well.

Table 6. Minor League Model Results

Variables	Model1	Model2
Relaxed Pitch Clock	0.052**	0.018
	(0.025)	(0.024)
Strict Pitch Clock	0.014	-0.042
	(0.060)	(0.054)
Log(Metropolitan Statistical Area Population)		0.96***
		(0.15)
Log(Win Percentage)		0.32***
		(0.04)
Previous Season Playoff Appearance		0.07***
		(0.01)
Intercept	8.99***	-3.80*
	(0.03)	(2.07)
Number of Observations	924	924
Adjusted R-Square	0.9582	0.9657
Overall Significance	2,259.69***	2,298.63***

Source: Sports Reference LLC. (2024), The Baseball Cube (2024), U.S. Census Bureau (2024), Statistics Canada (2024).
Notes: Robust Standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels respectively. This model contains Team and Year fixed effects.

VII. Conclusion

This analysis was conducted in order to answer the question of whether or not rule changes in Major League Baseball and minor league baseball have a significant impact on attendance numbers. The models show that the answer is actually league-dependent. At the Major League level, the 2023 rule changes had a positive, statistically significant impact on attendance. At the minor league level; however, the impact of the rule changes was effectively nothing. It is worth noting that the interpretation of these results should be done cautiously, particularly at the Major League level. Since the rule changes have only been implemented for one season so far, it was not possible to run a model with year fixed effects. While a year polynomial and a COVID-19 indicator variable are incorporated to alleviate part of this issue, the positive coefficient may still contain bias due to 2023 being the first year since the pandemic that all stadiums functioned at full capacity for the entire season.

How should Major League Baseball utilize these results? In essence, this analysis indicates that Major League Baseball should continue to explore rule changes largely utilizing the same methods they already do. Rule changes appear to have no statistically significant impact on attendance numbers, which make it perfect to use as a guinea pig. Utilizing their own data, the MLB can figure out which statistics or game factors correlate with attendance. Then, they can theorize rule changes that would address or improve these statistics and factors. Next, these changes can be tested at the minor league level, with little to no impact on the attendance numbers. If these changes affect gameplay in the desired manner, do not have negative impacts on player health, and receive positive reviews from players, they can then implement the changes at the Major League level. The player health aspect garners the most attention, especially because it may take time to notice its impact. Major League Baseball has experienced an increase in pitcher arm injuries since the introduction of the new rules, specifically those requiring Tommy John surgery. While no research has been done to signal causation, player health is certainly a concern that all rule changes should consider.

Future research on this topic consists of multiple different avenues, both specifically focused on the 2023 rule changes and otherwise. At a minimum, an analysis a few years from now with multiple years of post-treatment data would be ideal. A model of this style would be able to contain fixed effects, as the post-rule change period would be longer than a single year. Furthermore, this research would grant more insight in the post-pandemic era, meaning that the bias from COVID-19 would be minimized. Another interesting research point would involve breaking the Major League teams into groups based on market. Teams such as the Dodgers and

Yankees sell out nearly every game, so rule changes have no room to have a positive impact on attendance for them (they cannot go over stadium capacity). Perhaps an analysis on small to mid-market teams could result in a more accurate result on the true impact of the 2023 rule changes on attendance. Lastly, future rule changes could also be analyzed in a similar way. Minor league baseball is currently experimenting with robot umpires and the ability to challenge balls and strikes. If these rules ever make it to the Major League level, it may be interesting to see if the potential for more accurate umpiring would have any impact on attendance.

References

Ahn, S. C., & Lee, Y. H. (2014). Major League Baseball attendance: Long-term analysis using factor models. *Journal of Sports Economics*, *15*(5), 451-477.

Barajas, A., & Gasparetto, T. (2023). They are not all the same: Determinants of attendance across different sectors in a stadium. *Plos one*, *18*(7), e0289331.

Bradbury, J. C. (2020). Determinants of attendance in major league soccer. *Journal of Sport Management*, *34*(1), 53-63.

Denaux, Z. S., Denaux, D. A., & Yalcin, Y. (2011). Factors affecting attendance of major league baseball: revisited. *Atlantic Economic Journal*, *39*, 117-127.

Gitter, S. R., & Rhoads, T. A. (2010). Determinants of minor league baseball attendance. *Journal of Sports Economics*, *11*(6), 614-628.

Houghtaling, Zachary, "Analysis of Minor League Rule Changes Effect on Stolen Bases" (2022). *Williams Honors College, Honors Research Projects.* 1554.

Lim, N., & Pedersen, P. M. (2022). Examining determinants of sport event attendance: multilevel analysis of a major league baseball season. *Journal of Global Sport Management*, 7(1), 181-198.

Ormiston, R. (2014). Attendance effects of star pitchers in major league baseball. *Journal of Sports Economics*, *15*(4), 338-364.

Shaw, Tyler, "An Analysis of the Impacts of Pitch Clock Experience on Major League Baseball" (2024). *CMC Senior Theses.* 3477.

Smith, E. E., & Groetzinger, J. D. (2010). Do fans matter? The effect of attendance on the outcomes of major league baseball games. *Journal of Quantitative Analysis in Sports*, *6*(1).

Sports Reference LLC. Baseball-Reference.com - Major League Statistics and Information. https://www.baseball-reference.com/. April 30, 2024.

Statistics Canada. Table 17-10-0135-01 Population estimates, July 1, by census metropolitan area and census agglomeration, 2016 boundaries. April 30, 2024.

The Baseball Cube. (n.d.). Minor League Attendance - The Baseball Cube. The Baseball Cube. https://www.thebaseballcube.com/content/minor_attendance/. March 15, 2024.

U.S. Census Bureau, Resident Population in Metropolitan Statistical Areas (MSA), retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/, April 30, 2024.

Appendix: SAS Codes

```
libname AEData "~/my_shared_file_links/u47408605/Data"
access=readonly;
run;

/* Importing MLB/MiLB Data */

proc import datafile="/home/u63038880/MySAS/MLB and MiLB Data.xlsx"
	out=work.AllData
	dbms=xlsx
	replace;
	range="AllData$A1:M1225";
	getnames=yes;
run;

/* Setting Data for MLB Models */

Data MLBModelData;
	set AllData;
	where Level = 0 and Year > 2011;
	if Year = 2021 then Covid = 1;
	else Covid = 0;
	if Year = 2023 then RuleChange = 1;
	else RuleChange = 0;
	if Team="Atlanta" and Year=2017 then NewBP = 1;
	else if Team="Miami" and Year=2012 then NewBP=1;
	else if Team="Texas" and Year=2021 then NewBP=1;
```

```
        else NewBP = 0;
        LogAvgAtt = log(AvgAtt);
        LogPayroll = log(Payroll);
        LogMSAPop = log(MSAPop);
        LogWinPct = log(WinPct);
        YearB = Year-2020;
        YearB2 = YearB * YearB;
        YearB3 = YearB2 * YearB;
run;

proc means data= mlbmodeldata;
run;

/* Major League Model 1: No Controls (except COVID) */

ods output ParameterEstimates=PEforModel1  DataSummary=DS1 Effects=E1
FitStatistics=FS1;
Proc SurveyReg Data=MLBModelData plots=none;
        Class Team /ref=first;
        Model LogAvgAtt = RuleChange Covid Team YearB YearB2 YearB3 /Solution AdjRsq;
run;
ods output close;

/* Major League Model 2: All Controls */

ods output ParameterEstimates=PEforModel2  DataSummary=DS2 Effects=E2
FitStatistics=FS2;
Proc SurveyReg Data=MLBModelData plots=none;
        Class Team /ref=first;
```

```
Model LogAvgAtt = RuleChange Covid LogPayroll LogWinPct LogMSAPop
PlayoffLYear NewBP Team YearB YearB2 YearB3 /Solution AdjRsq;

run;

ods output close;


/* Major League Model 3: All Controls Except Population (Was insignificant in Model 2) */


ods output ParameterEstimates=PEforModel3  DataSummary=DS3 Effects=E3
FitStatistics=FS3;

Proc SurveyReg Data=MLBModelData plots=none;

Class Team /ref=first;

Model LogAvgAtt = RuleChange Covid LogWinPct LogPayroll PlayoffLYear NewBP
Team YearB YearB2 YearB3 /Solution AdjRsq;

run;

ods output close;


/* Cleaning up Regression Results */


Data Table_Long;

length Model $10; /* Makes sure the variable Model has the right length and its values
are not truncated */

length Parameter $30; /* Makes sure the variable Parameter has the right length and its
values are not truncated */

set PEforModel1 PEforModel2 PEforModel3 indsname=M; /*"indsname" creates an
indicator variable (here I call it "M") that tracks the name of databases use in the "set" statement
*/

where Estimate ne 0

and substr(Parameter,1,4) ne "Team"

and substr(Parameter,1,5) ne "Year";

keep Parameter Model EditedResults;

if      M="WORK.PEFORMODEL1" then Model="Model1";
```

```
	else if M="WORK.PEFORMODEL2" then Model="Model2";
	else if M="WORK.PEFORMODEL3" then Model="Model3";

	Length Star $3;
	If Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";
		else Star="";

	EditedResults=cats(put(Estimate, comma16.2),Star);
	output;

	EditedResults=Cats("(",put(StdErr, comma16.2),")");
	output;

run;

/* Sorting Regression Model Results */

proc sort data=Table_Long out=Table_Long_Sorted;
	by Model Parameter;
run;

/*Creating separate results columns (in the form of separate databases) corresponding to each
model */

data Model1Results(rename=(EditedResults=Model1))
	 Model2Results(rename=(EditedResults=Model2))
	 Model3Results(rename=(EditedResults=Model3));
```

```
	set Table_Long_Sorted;
	drop Model;
	if Model="Model1" then output Model1Results;
	else if Model="Model2" then output Model2Results;
	else if Model="Model3" then output Model3Results;
run;

/* Creating final results table */

data Table_Wide;
	merge Model1Results Model2Results Model3Results;
	by Parameter;

	length Order 3;
	if Parameter = "RuleChange" then Order = 1;
		else if Parameter = "Intercept" then Order = 3;
		else Order = 2;



	if mod(_n_,2) = 1 then Variables=Parameter;


run;

proc sort data=Table_Wide;
	by Order;
run;

proc format;
```

```
        value $VariableName(default=50)
        "RuleChange"="Rule Changes"
        "Covid"="Post-Covid Year (2021)"
        "LogMSAPop"="Log(Metropolitan Statistical Area Population)"
        "LogPayroll"="Log(Payroll)"
        "PlayoffLYear"="Previous Season Playoff Appearance"
        "WinPct"="Win Percentage"
        "NewBP"="New Ballpark";
run;


/* Creating the rows for other statistics */

/* The row for the number of observations */

Data NumofObs;
        merge DS1(rename=(nValue1=NVModel1) drop=CValue1)
                DS2(rename=(nValue1=NVModel2) drop=CValue1)
                DS3(rename=(nValue1=NVModel3) drop=CValue1);
        where Label1="Number of Observations";
        Model1= put(NVModel1,comma16.);
        Model2= put(NVModel2,comma16.);
        Model3= put(NVModel3,comma16.);
        keep Label1 Model1 Model2 Model3;
run;

/* The row for Adjusted R Squared */
```

```
Data AdjRSquared;
	merge FS1(rename=(CValue1=Model1) drop=nvalue1)
		FS2(rename=(CValue1=Model2) drop=nvalue1)
		FS3(rename=(CValue1=Model3) drop=nvalue1);
	where Label1="Adjusted R-Square";
run;

Data OSM1(rename=(FValueNew=Model1))
	 OSM2(rename=(FValueNew=Model2))
	 OSM3(rename=(FValueNew=Model3));
	set E1 E2 E3 indsname=M;
	where Effect="Model";
	If ProbF le 0.01 then Star="***";
		else if ProbF le 0.05 then Star="**";
		else if ProbF le 0.1 then Star="*";
		else Star="";
	Label1="Overall Significance";
	FValueNew=cats(put(Fvalue,Comma16.2),Star);
	if M="WORK.E1" then output OSM1;
	else if M="WORK.E2" then output OSM2;
	else if M="WORK.E3" then output OSM3;
	keep Label1 FValueNew;
run;

Data OverallSig;
	merge OSM1 OSM2 OSM3;
	by Label1;
run;
```

```
/* Combine all rows for other statistics */

data OtherStat;
	set NumofObs AdjRSquared OverallSig;
	rename Label1=Variables;
run;

data FinalMerge;
	set Table_Wide OtherStat;
run;

/* Print the clean results table */

ods excel file="/home/u63038880/MySAS/Senior Project MLB Model.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */
Title "Major League Model Results";
footnote justify=left "Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics
Canada (2024).";
proc print data=FinalMerge noobs;
	var Variables;
	format Variables $VariableName.;
	var Model1 Model2 Model3 / style(header)={Just=Center} style(data)={Just=Center
tagattr="type:string"};
run;
ods excel close;

/* Setting Minor League Model Data */
```

```
Data MiLBModelData;
	set AllData;
	where Year < 2023 and WinPct ne 0;
	Treatment = Level;
	if Level = 0 then RPitchClock = 0;
	else if Year < 2015 then RPitchClock = 0;
	else if Year > 2014 and Year < 2022 then RPitchClock = 1;
	else if Year = 2022 then RPitchClock = 0;
	if Level = 0 then SPitchClock = 0;
	else if Year ne 2022 then SPitchClock = 0;
	else if Year = 2022 then SPitchClock = 1;
	LogAvgAtt = log(AvgAtt);
	LogMSAPop = log(MSAPop);
	LogWinPct = log(WinPct);
	YearB = Year-2020;
	YearB2 = YearB * YearB;
	YearB3 = YearB2 * YearB;
run;

proc means data= MiLBModelData;
	class Level;
run;

/* Balance of Regressors Test */

Proc TTest Data = MiLBModelData;
	Where Year < 2015;
```

```
	Var WinPct MSAPop PlayoffLYear;
	Class Treatment;
run;

/* Parallel Trend Statistical Test */

/* Parallel Trend Test 1: No Controls */

ods output ParameterEstimates=PEforPT1  DataSummary=DSPT1 Effects=EPT1
FitStatistics=FSPT1;
proc surveyreg Data=MiLBModelData;
	Where Year<2015;
	Model LogAvgAtt = Treatment YearB YearB2 YearB3 Treatment*YearB
Treatment*YearB2 Treatment*YearB3 /Solution AdjRsq;
run;
ods output close;

/* Parallel Trend Test 2: With Controls */

ods output ParameterEstimates=PEforPT2  DataSummary=DSPT2 Effects=EPT2
FitStatistics=FSPT2;
proc surveyreg Data=MiLBModelData;
	Where Year<2015;
	Model LogAvgAtt = Treatment LogWinPct LogMSAPop PlayoffLYear YearB YearB2
YearB3 Treatment*YearB Treatment*YearB2 Treatment*YearB3 /Solution AdjRsq;
run;
ods output close;

/* Cleaning up Parallel Trend Results */
```

```
Data Table_Long;
	length Model $10; /* Makes sure the variable Model has the right length and its values are not truncated */
	length Parameter $30; /* Makes sure the variable Parameter has the right length and its values are not truncated */
	set PEforPT1 PEforPT2 indsname=M; /*"indsname" creates an indicator variable (here I call it "M") that tracks the name of databases use in the "set" statement */
	where Estimate ne 0
		and substr(Parameter,1,4) ne "Team"
		and substr(Parameter,1,5) ne "Year";
	keep Parameter Model EditedResults;
	if	M="WORK.PEFORPT1" then Model="Model1";
	else if M="WORK.PEFORPT2" then Model="Model2";

	Length Star $3;
	If Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";
		else Star="";

	EditedResults=cats(put(Estimate, comma16.2),Star);
	output;

	EditedResults=Cats("(",put(StdErr, comma16.2),")");
	output;

run;

/* Sorting Regression Model Results */
```

```
proc sort data=Table_Long out=Table_Long_Sorted;
	by Model Parameter;
run;

/*Creating separate results columns (in the form of separate databases) corresponding to each
model */

data PT1Results(rename=(EditedResults=Model1))
	 PT2Results(rename=(EditedResults=Model2));
	set Table_Long_Sorted;
	drop Model;
	if Model="Model1" then output PT1Results;
	else if Model="Model2" then output PT2Results;
run;

/* Creating final results table */

data Table_Wide;
	merge PT1Results PT2Results;
	by Parameter;

	length Order 3;
	if Parameter = "Treatment" then Order = 1;
		else if substr(Parameter,1,5) = "YearB" then Order = 3;
		else if substr(Parameter,1,10) = "Treatment*" then Order = 4;
		else if Parameter = "Intercept" then Order = 5;
		else Order = 2;
```

```
	if mod(_n_,2) = 1 then Variables=Parameter;

run;

proc sort data=Table_Wide;
	by Order;
run;

proc format;
	value $VariableName(default=50)
	"LogMSAPop"="Log(Metropolitan Statistical Area Population)"
	"LogWinPct"="Log(Win Percentage)"
	"PlayoffLYear"="Previous Season Playoff Appearance"
	"WinPct"="Win Percentage"
	"NewBP"="New Ballpark";
run;



/* Creating the rows for other statistics */

/* The row for the number of observations */

Data NumofObs;
	merge DSPT1(rename=(nValue1=NVModel1) drop=CValue1)
		DSPT2(rename=(nValue1=NVModel2) drop=CValue1);
	where Label1="Number of Observations";
	Model1= put(NVModel1,comma16.);
```

```
	Model2= put(NVModel2,comma16.);;
	keep Label1 Model1 Model2;
run;

/* The row for Adjusted R Squared */

Data AdjRSquared;
	merge FSPT1(rename=(CValue1=Model1) drop=nvalue1)
	      FSPT2(rename=(CValue1=Model2) drop=nvalue1);
	where Label1="Adjusted R-Square";
run;

Data OSMPT1(rename=(FValueNew=Model1))
	 OSMPT2(rename=(FValueNew=Model2));
	set EPT1 EPT2 indsname=M;
	where Effect="Model";
	If ProbF le 0.01 then Star="***";
		else if ProbF le 0.05 then Star="**";
		else if ProbF le 0.1 then Star="*";
		else Star="";
	Label1="Overall Significance";
	FValueNew=cats(put(Fvalue,Comma16.2),Star);
	if M="WORK.EPT1" then output OSMPT1;
	else if M="WORK.EPT2" then output OSMPT2;
	keep Label1 FValueNew;
run;

Data OverallSig;
```

```
	merge OSMPT1 OSMPT2;
	by Label1;
run;

/* Combine all rows for other statistics */

data OtherStat;
	set NumofObs AdjRSquared OverallSig;
	rename Label1=Variables;
run;

data FinalMerge;
	set Table_Wide OtherStat;
run;


/* Print the clean results table */

ods excel file="/home/u63038880/MySAS/Senior Project Parallel Trend Test.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */
Title "Parallel Trend Statistical Test";
footnote justify=left "Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics
Canada (2024).";
proc print data=FinalMerge noobs;
	var Variables;
	format Variables $VariableName.;
	var Model1 Model2 / style(header)={Just=Center} style(data)={Just=Center
tagattr="type:string"};
run;
```

```
ods excel close;


/* Minor League Model 1: No Controls */


ods output ParameterEstimates=PEforModel1  DataSummary=DS1 Effects=E1
FitStatistics=FS1;

Proc SurveyReg Data=MiLBModelData plots=none;

	Class Team YearB /ref=first;

	Model LogAvgAtt = RPitchClock SPitchClock Team YearB /Solution AdjRsq;

run;

ods output close;


/* Minor League Model 2: All Controls */


ods output ParameterEstimates=PEforModel2  DataSummary=DS2 Effects=E2
FitStatistics=FS2;

Proc SurveyReg Data=MiLBModelData plots=none;

	Class Team YearB /ref=first;

	Model LogAvgAtt = RPitchClock SPitchClock LogWinPct LogMSAPop PlayoffLYear
Team YearB /Solution AdjRsq;

run;

ods output close;


/* Cleaning up Regression Results */


Data Table_Long;

	length Model $10; /* Makes sure the variable Model has the right length and its values
are not truncated */

	length Parameter $30; /* Makes sure the variable Parameter has the right length and its
values are not truncated */
```

```
	set PEforModel1 PEforModel2 indsname=M; /*"indsname" creates an indicator variable (here I call it "M") that tracks the name of databases use in the "set" statement */
	where Estimate ne 0
		and substr(Parameter,1,4) ne "Team"
		and substr(Parameter,1,5) ne "Year";
	keep Parameter Model EditedResults;
	if	M="WORK.PEFORMODEL1" then Model="Model1";
	else if M="WORK.PEFORMODEL2" then Model="Model2";

	Length Star $3;
	If Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";
		else Star="";

	EditedResults=cats(put(Estimate, comma16.2),Star);
	output;

	EditedResults=Cats("(",put(StdErr, comma16.2),")");
	output;

run;

/* Sorting Regression Model Results */

proc sort data=Table_Long out=Table_Long_Sorted;
	by Model Parameter;
run;
```

```
/*Creating separate results columns (in the form of separate databases) corresponding to each
model */

data Model1Results(rename=(EditedResults=Model1))
            Model2Results(rename=(EditedResults=Model2));
        set Table_Long_Sorted;
        drop Model;
        if Model="Model1" then output Model1Results;
        else if Model="Model2" then output Model2Results;
run;

/* Creating final results table */

data Table_Wide;
        merge Model1Results Model2Results;
        by Parameter;

        length Order 3;
        if Parameter = "RPitchClock" then Order = 1;
                else if Parameter = "SPitchClock" then Order = 2;
                else if Parameter = "Intercept" then Order = 4;
                else Order = 3;



        if mod(_n_,2) = 1 then Variables=Parameter;

run;

proc sort data=Table_Wide;
```

```
    by Order;
run;

proc format;
    value $VariableName(default=50)
    "DID"="Rule Changes"
    "Covid"="Post-Covid Year (2021)"
    "LogMSAPop"="Log(Metropolitan Statistical Area Population)"
    "LogPayroll"="Log(Payroll)"
    "PlayoffLYear"="Previous Season Playoff Appearance"
    "WinPct"="Win Percentage"
    "NewBP"="New Ballpark";
run;


/* Creating the rows for other statistics */

/* The row for the number of observations */

Data NumofObs;
    merge DS1(rename=(nValue1=NVModel1) drop=CValue1)
          DS2(rename=(nValue1=NVModel2) drop=CValue1);
    where Label1="Number of Observations";
    Model1= put(NVModel1,comma16.);
    Model2= put(NVModel2,comma16.);;
    keep Label1 Model1 Model2;
run;
```

```
/* The row for Adjusted R Squared */

Data AdjRSquared;
        merge FS1(rename=(CValue1=Model1) drop=nvalue1)
                FS2(rename=(CValue1=Model2) drop=nvalue1);
        where Label1="Adjusted R-Square";
run;

Data OSM1(rename=(FValueNew=Model1))
         OSM2(rename=(FValueNew=Model2));
        set E1 E2 E3 indsname=M;
        where Effect="Model";
        If ProbF le 0.01 then Star="***";
                else if ProbF le 0.05 then Star="**";
                else if ProbF le 0.1 then Star="*";
                else Star="";
        Label1="Overall Significance";
        FValueNew=cats(put(Fvalue,Comma16.2),Star);
        if M="WORK.E1" then output OSM1;
        else if M="WORK.E2" then output OSM2;
        keep Label1 FValueNew;
run;

Data OverallSig;
        merge OSM1 OSM2;
        by Label1;
run;
```

```
/* Combine all rows for other statistics */

data OtherStat;
	set NumofObs AdjRSquared OverallSig;
	rename Label1=Variables;
run;

data FinalMerge;
	set Table_Wide OtherStat;
run;


/* Print the clean results table */

ods excel file="/home/u63038880/MySAS/Senior Project MiLB Model.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */
Title "Minor League Model Results";
footnote justify=left "Sports Reference LLC. (2024), U.S. Census Bureau (2024), Statistics
Canada (2024).";
proc print data=FinalMerge noobs;
	var Variables;
	format Variables $VariableName.;
	var Model1 Model2 / style(header)={Just=Center} style(data)={Just=Center
tagattr="type:string"};
run;
ods excel close;
```