Winning Over the Fans:
The Impact of Team Performance on MLB Attendance

Tyler Christianson, Sean Faikish, Liam Smith
University of Notre Dame
MSBR 70320: Time Series Forecasting

Introduction

Professional sports is a business. Team success plays a crucial role in generating money for these businesses, whether through merchandising, television agreements, or ticket sales. Across all sports, it is no surprise that consistently winning teams tend to generate more off-field success. Winning attracts more fans, better players, and lucrative sponsorships that will help continue franchise growth. However, sustaining franchise growth has been a challenge for Major League Baseball in recent years. Several factors, such as the absence of a salary cap, the reluctance of owners to spend, and the widening gap between small and large market teams, have contributed to these challenges. These factors not only affect a team's season performance but also affect the level of fan engagement.

Often referred to as "America's Greatest Pastime," Major League Baseball (MLB) has a rich history of tradition, culture, and fan loyalty.[1] Baseball brings people together, but how much does winning truly impact an MLB team's attendance? Our project will examine how an MLB team's on-field performance impacts their attendance over time. Ultimately, the following key questions will drive our analysis:

1. Should the MLB be incentivized to pursue winning to increase attendance and increase profitability?
2. How strongly does a team's performance (e.g., winning percentage) influence attendance trends?
3. Which forecasting models provide the best predictive accuracy for an MLB team's attendance?

To answer these questions, we used historical attendance data and historical team records across the MLB and applied time series forecasting techniques to analyze trends. After checking for Random Walk, we determined the most effective time series models for three MLB teams to forecast future attendance records. Understanding the relationships can help MLB teams, stadium partners, and sports analysts optimize resource allocation, ticket pricing, and marketing strategies to maximize the two best things that come out of sports: fan engagement and revenue. Additionally, this analysis aims to provide valuable recommendations for executives, owners, and fans across the MLB looking to make a significant impact on their franchise.

Data Methodology

Data Selection

Our final dataset compiled scraped data from two main data frames on BaseballReference.com: MLB Standings and MLB Attendance. Both datasets include observations from 1903 through the completion of the 2024 season. Data on the national Consumer Price Index was also scraped to perform an overarching league analysis. While these findings were not significant, they will be included for continuity's sake.

---

[1] https://baseballhall.org/baseball-history-american-history-and-you

MLB Standings variables:

- Historical team records
- Win/loss statistics
- Average runs scored/allowed
- Strength of Schedule

MLB Attendance variables:

- Total season attendance
- Attendance per game
- Average pitcher/batter age
- Number of Hall of Famers

CPI variables:

- Monthly CPI from 1913-2024
- Change in CPI over last year
- Average CPI for all time

Several preprocessing steps were applied to ensure data accuracy, including cleaning and removing, and imputing missing values where necessary. Notably, the MLB did not play in the 2020 season due to the Covid-19 pandemic. Our initial analysis excluded that season, but to maintain continuity, missing wins and average attendance values were imputed by calculating the average from the previous (2019) and following (2021) seasons.

Data Processing

As mentioned above, the final dataset included attendance records for all MLB teams spanning more than a century. We divided the data into these key sections:

- League-wide trends: Aggregated data capturing total MLB attendance over the years.
- Team-specific analysis: Focused our case studies on the Chicago White Sox, Detroit Tigers, and Pittsburgh Pirates. (Team observations spanned across a 120-year window and were not a random walk during autocorrelation testing)
- Economic adjustments: Incorporated the Customer Price Index (CPI) as an exogenous variable to account for the evolving financial landscape of consumers and the impact it has on league-wide MLB attendance.

Data Partition

Our datasets included the yearly attendance average for each MLB team from 1903 to 2024. We split the data into training and validation periods. Initially, our training period spanned from 1903 to 2019, but due to the imputation of the 2020 season (we did not want the COVID-19 season to affect our validation set), we extended the training period to 2021. The validation period spanned the last three years, from 2022 to 2024. Ultimately, we combined the training and validation

periods to help build our forecasts. Exhibit 1 illustrates our Full League model's training and validation sets, serving as an example for the other three MLB teams.

Model Feasibility

Checking for Random Walk

After data selection and processing, we tested for random walks on both full-league and individual team attendance. As shown in Exhibit 2, our ACF results confirmed that league-wide attendance follows a random walk, signifying the naive models are likely the best forecasting method.

However, individual teams that span the entire 120-year history of the MLB, like the White Sox, Tigers, and Pirates, exhibited more predictable patterns shown in Exhibits 3, 4, and 5, respectively. The team-focused ACF plots move up/down, whereas the Full League plot shows a steady downward trend at each lag.

We conducted stationarity tests to determine the suitability of different forecasting models. The testing provided the following t-statistics and p-values for the teams.

Team	T-Statistic	P-Value
Chicago White Sox	-2.47	0.013
Detroit Tigers	-2.71	0.007
Pittsburgh Pirates	-2.94	0.003

Model Selection

Forecasting models tested for league-wide analysis:
- Naive model
- ARIMAX with CPI
- ARIMAX with lagged CPI

Forecasting models tested for team-specific analysis:
- Naive Model
- Linear Trend Model
- Polynomial Trend Model
- Moving Average Model
- Holt Model
- Auto ARIMA
- ARIMAX with wins
- ARIMAX with lagged wins

The final model selection was done based on Mean Absolute Percentage Error (MAPE), with the best model chosen for the league and each team (Exhibit 6 and Exhibit 7).

<u>Model Forecasting</u>

Full League (Exhibit 8): Despite the full league model being a random walk, we determined that it would be worthwhile to check if CPI had any influence on attendance trends. Given that CPI serves as a proxy for inflation and overall economic conditions in the US, examining its impact on league-wide attendance could provide insights into how spending power impacts fan behavior and ticket sales all the time. After testing naive and ARIMA models, we found that the best approach for modeling league-wide attendance was an ARIMAX model with lagged CPI as a predictor. This suggests that recent economic conditions impact attendance patterns.

Chicago White Sox (Exhibit 9): For the Chicago White Sox, we selected an ARIMAX model with current-season wins as a predictor as the best for modeling. This approach allows us to incorporate both the time series component captured by a normal ARIMA model as well as the direct influence of team performance in any given season. By incorporating current wins into the model, we are taking into account how White Sox fans are actively changing in behavior based on their team's success. According to this, the more wins, the more fans will come to games, and the less wins, the quicker these fans will stop coming to games. This model provides a dynamic view of attendance patterns and shows that fan interest is often tied to team performance.

Detroit Tigers (Exhibit 10): For the Detroit Tigers, we selected a Moving Average (MA) model as the best approach for forecasting attendance. This model does well at capturing short-term fluctuations and smoothing out noise. The MA model's strength lies in its ability to quickly adapt to shifts in attendance patterns, making it well-suited for short-term forecasting problems like this one. This model performing the best tells us that fans of the Tigers are willing to stick with their team, even if they fall for a few seasons in terms of success. If their wins stay low, however, fans will respond.

Pittsburgh Pirates (Exhibit 11): For the Pittsburgh Pirates, a Moving Average model was once again the best approach for forecasting attendance. Similarly to Tigers' fans, Pirates' fans may take a few years to respond to fluctuations in their team's performance. Their attendance responds a bit slower to fluctuations in success as opposed to the White Sox.

<u>Model Limitations</u>

While the ARIMAX and Rolling Moving Average models demonstrated strong forecasting ability, each has their limitations. The current wins ARIMAX model assumes a linear and immediate correlation between team performance and attendance. Yet, this model does not account for lagged effects, which means historical momentum or team reputation over a long period is not properly applied. In addition, ARIMAX models are susceptible to overfitting if there are numerous exogenous variables, which could reduce generalizability when predicting future seasons; however, only one variable was included in this model.

Although the Rolling Moving Average model is effective in separating short-run attendance trends, it lacks the ability to incorporate outside predictors such as team performance, economic conditions, or previous franchise prestige. This model assumes that current trends will continue and, therefore, is less suitable for teams with wild volatility in attendance due to matters that occur in the league, such as surprise playoff streaks, stadium construction, or front-office transitions, just to name a few. Moreover, it will not likely capture sudden shifts in fan reaction due to team identity, high player signings, or market factors.

Conclusion

Our analysis confirms that wins are a crucial driver of MLB attendance. The White Sox experienced a strong correlation between recent wins and attendance, with the ARIMAX model included with wins as the best-performing model. This implies that successful short-term teams will record a higher rate of attendance because the fans are more likely to support a winning team within the current season.

The Tigers and Pirates, on the other hand, showed attendance behaviors best modeled by a Rolling Moving Average, and their attendance behavior reacted less immediately to performance change. This would mean that winning matters, but other issues, such as historical performance, team brand, fans' loyalty, and market situation, drive the attendance levels of the 120-year teams.

Applications

Looking at the broader landscape of the MLB, the results from our analysis can help drive decision-making at three key levels: Executive, Ownership, and Fan. Each plays an important role in driving team success.

Executive: MLB Executives should expect overall league attendance to fluctuate based on purchasing power, as our analysis shows that lagged CPI is a good predictor of league attendance. This indicates that economic conditions can influence consumer buying behavior (i.e. decision to buy game tickets). Understanding this relationship can help executives make better decisions on promotions, broadcast rights, and ticket pricing. Additionally, the MLB should look into implementing a salary cap to maintain a competitive balance across the league. Without a salary cap, small market teams lose out on high-impact players. Small market teams like the Athletics (~$50M payroll) stand little chance against "MLB Superteams" like the Dodgers, who have payrolls exceeding $1B.[2] A salary cap will diversify the amount of superstar players across teams, resulting in a more league-wide competitive balance. This will hopefully lead to more wins for small market teams, and ultimately drive more fan engagement.

Ownership: Similar to executives, owners should use this analysis to understand the impact fans have on the growth of their franchise. One effective way to continue to maintain/grow attendance is by consistent investments. This includes investments in facilities, game-day personnel, fan experiences, and most importantly, in team performance. Owners must invest in their teams and increase their payroll spending. In recent years, the formula for more wins has become clear: higher payroll leads to better players, better players leads to a better team, and a better team

---

[2] https://www.spotrac.com/mlb/payroll/_/year/2025/sort/cap_total2

results in more wins. Our analysis supports the idea that wins are a good predictor of attendance, meaning more wins leads to better attendance.

Fan: Most fans believe their actions will go unnoticed. In the grand scheme of things, they are often right. One (or a few) frustrated fans will not catch the attention of many MLB owners and executives. However, our analysis should offer a sense of hope. Fans have a powerful voice that can drive real change for management. Change may not come from one individual, but through collective efforts, they can make a real impact. The simplest recommendation we can give based on our analysis is this: if you are frustrated with your team, do not go to games. Fanbases like the Oakland Athletics followed this strategy, with attendance boycotts and vocalizing fan frustration. Their lack of spending, fan engagement, and attendance has resulted in the decision to relocate to a new city. As a fan, your voice, your actions, and your loyalty have an impact that can reach team owners and executives.

<u>Future work</u>

Although our analysis offers effective forecasts, further improvements could enhance the accuracy of our models. Given more time, we would introduce postseason performance as a new predictor to capture the impact a playoff run has on fan attendance (i.e. 0 for missed playoffs, 1 for WC, 2 for Divisional, 3 for Championship, 4 for World Series). Additionally, we would look to add additional exogenous variables to track other economic conditions. Finally, instead of looking at attendance from a yearly perspective, we would look to get more granular by analyzing a team's game-to-game attendance fluctuations, capturing seasonal patterns and short-term changes more effectively.

Appendix

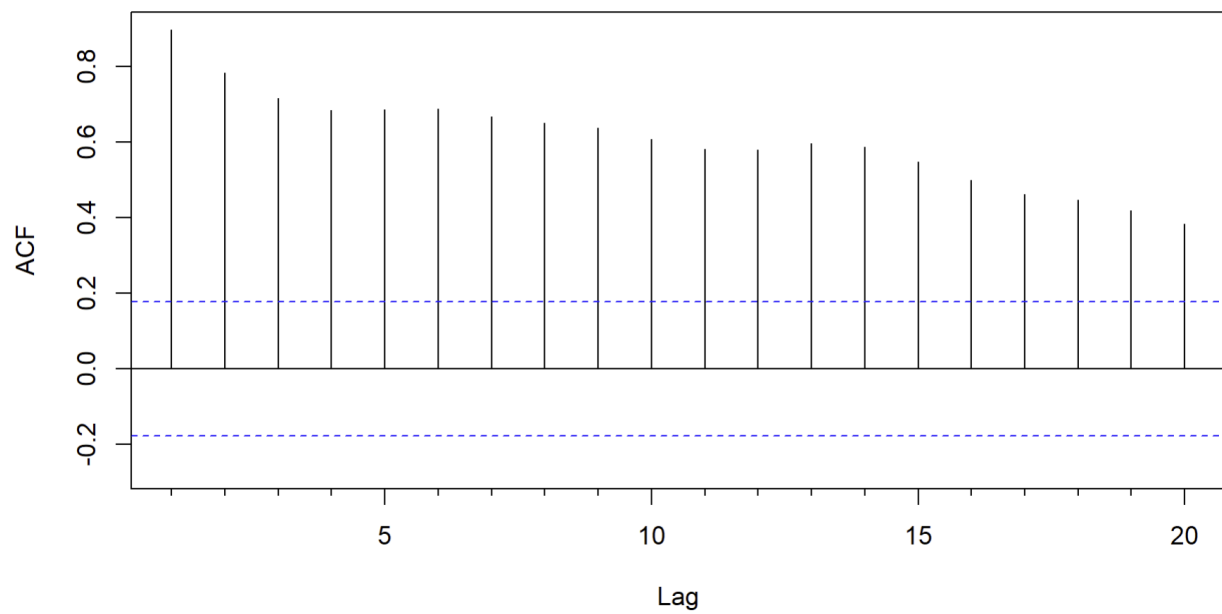## Exhibit 1: Training/Validation Split

Full League Attendance
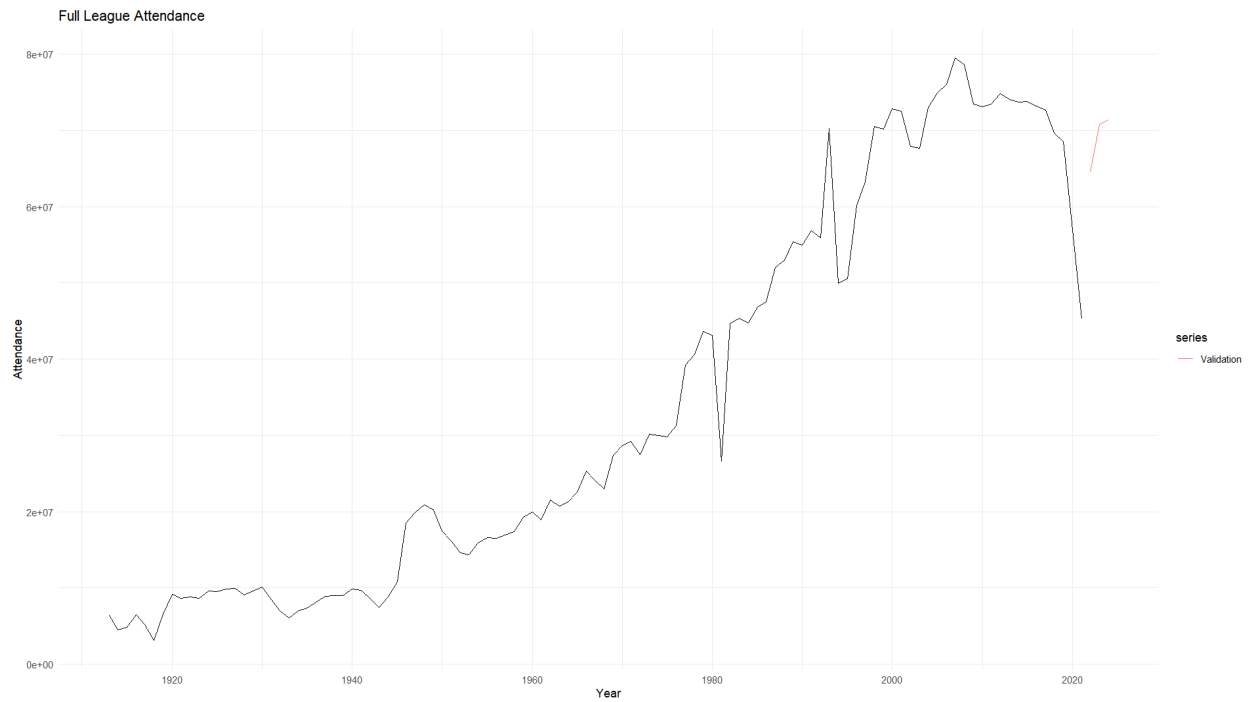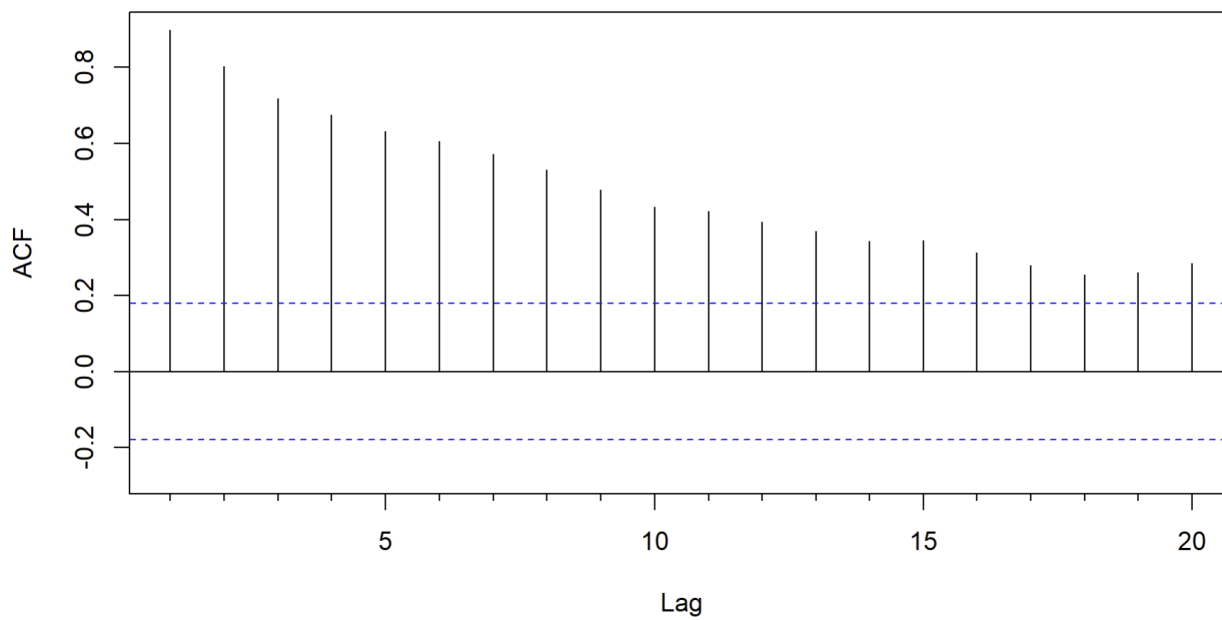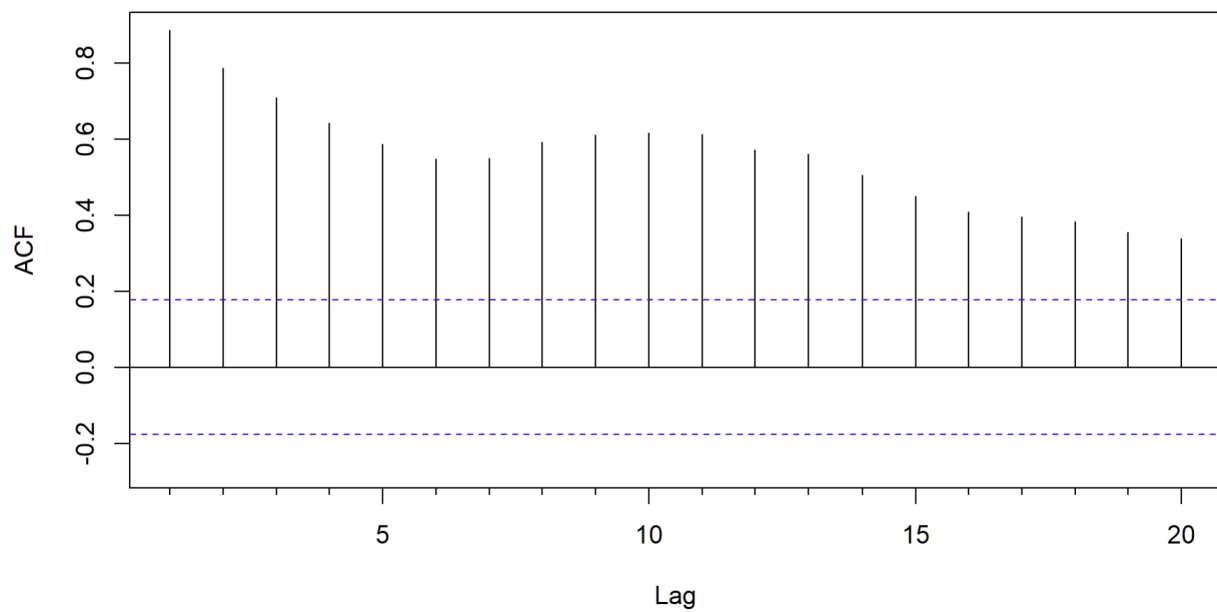
Exhibit 2: Full League ACF
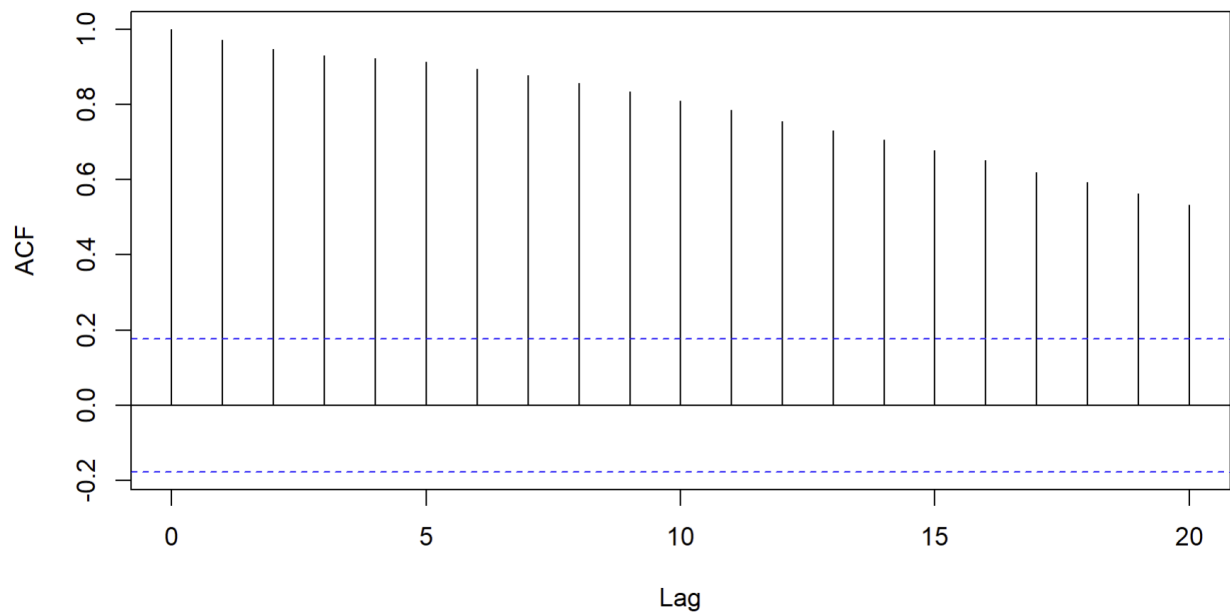
Exhibit 3: Chicago White Sox ACF

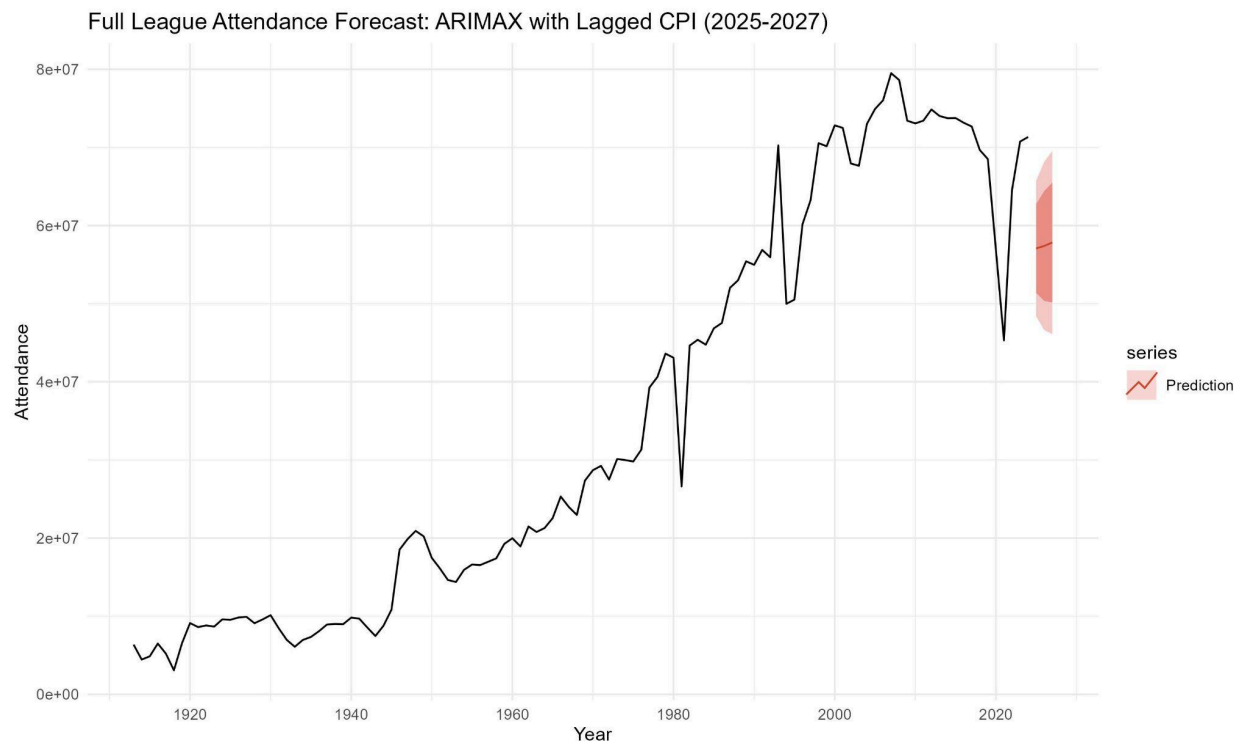**Exhibit 4: Detroit Tigers ACF**

Exhibit 5: Pittsburgh Pirates ACF

Exhibit 6: Full League MAPE

Model	MAPE
Naive	34.09
ARIMAX w/ CPI	38.04
ARIMAX w/ Lagged CPI	32.14

Exhibit 7: Team Specific MAPE

	White Sox	Tigers	Pirates
Model	**MAPE**	**MAPE**	**MAPE**
Naive	13.91	34.11	42.99
Linear Trend	27.68	42.79	27.20
Polynomial Trend	37.22	40.66	31.46
Moving Average	13.58	6.85	12.89
Holt	14.11	29.72	39.93
Auto ARIMA	13.91	34.11	42.99
ARIMAX w/ Wins	13.95	34.85	35.18
ARIMAX w/ Lagged Wins	13.32	34.08	42.92

Exhibit 8: Full League Model Predictions

Full League Attendance Forecast: ARIMAX with Lagged CPI (2025-2027)

Exhibit 9: White Sox Model Predictions

White Sox Attendance Forecast: SARIMAX with Lagged Wins (2025-2027)
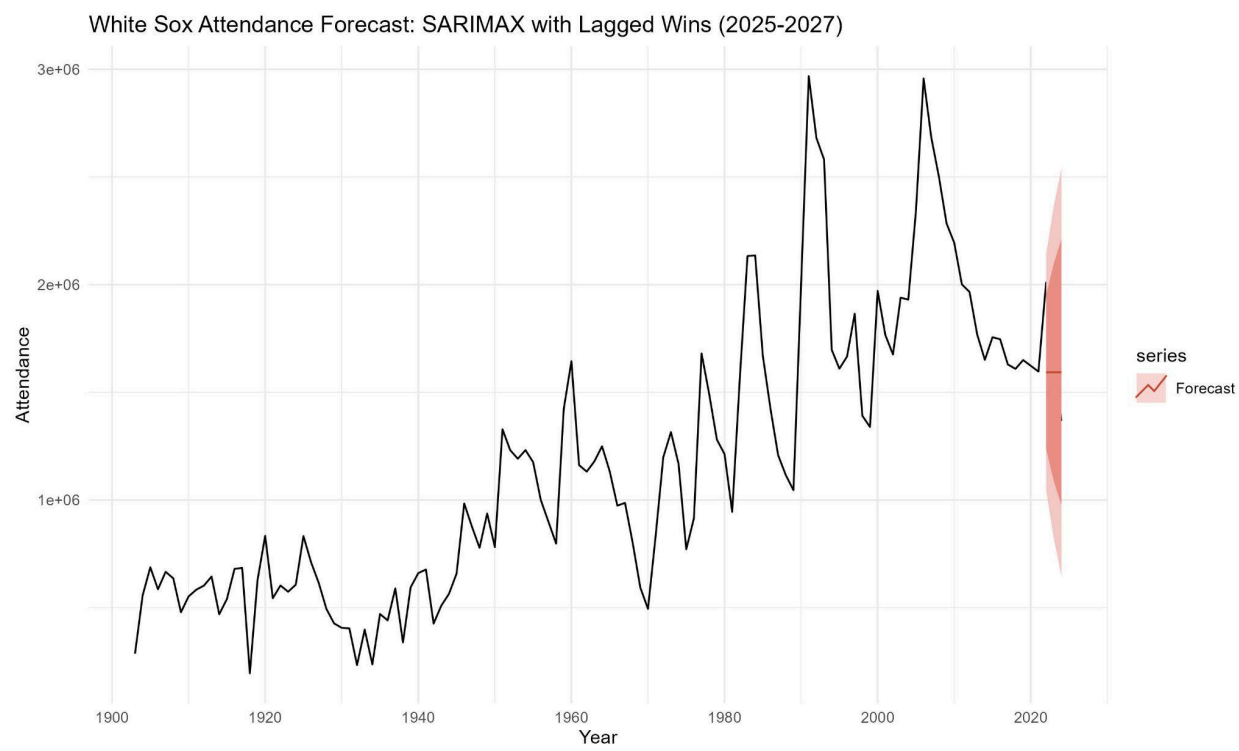
Exhibit 10: Detroit Tigers Model Predictions

Tigers Attendance Forecast: Moving Average Method (2025-2027)
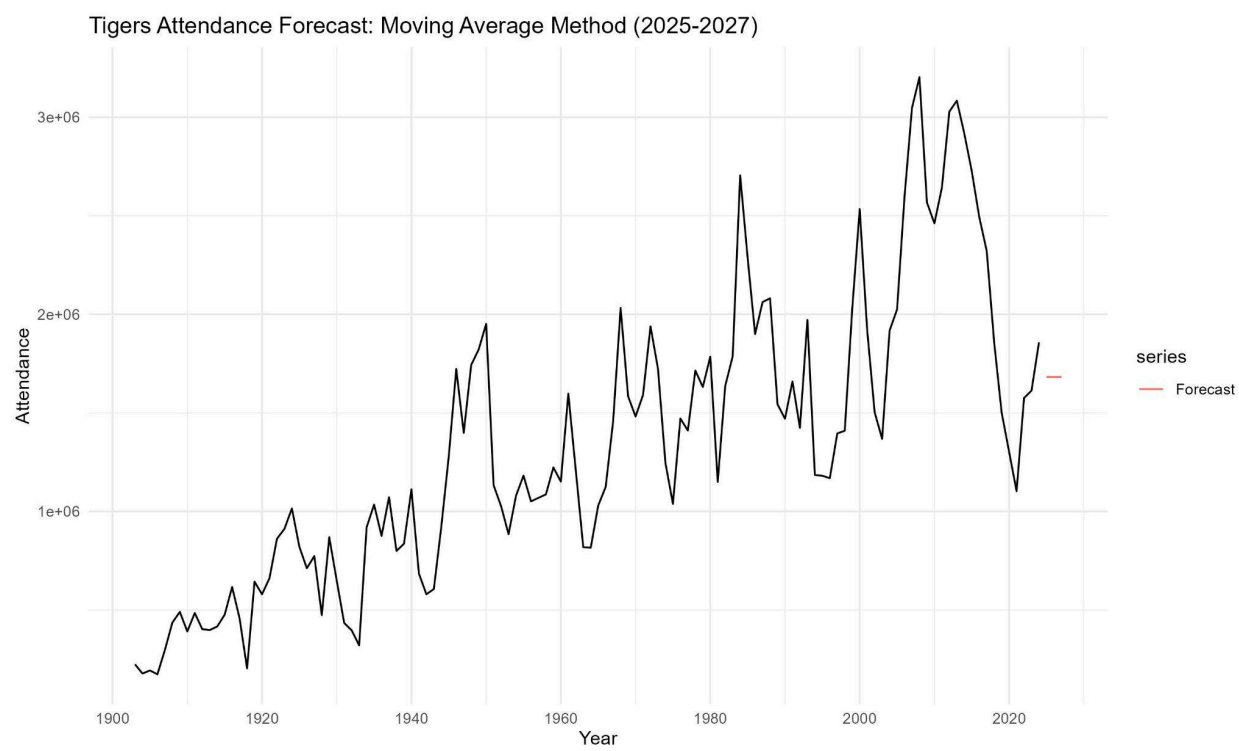
Exhibit 11: Pittsburgh Pirates Model Predictions

Pirates Attendance Forecast: Moving Average Method (2025-2027)
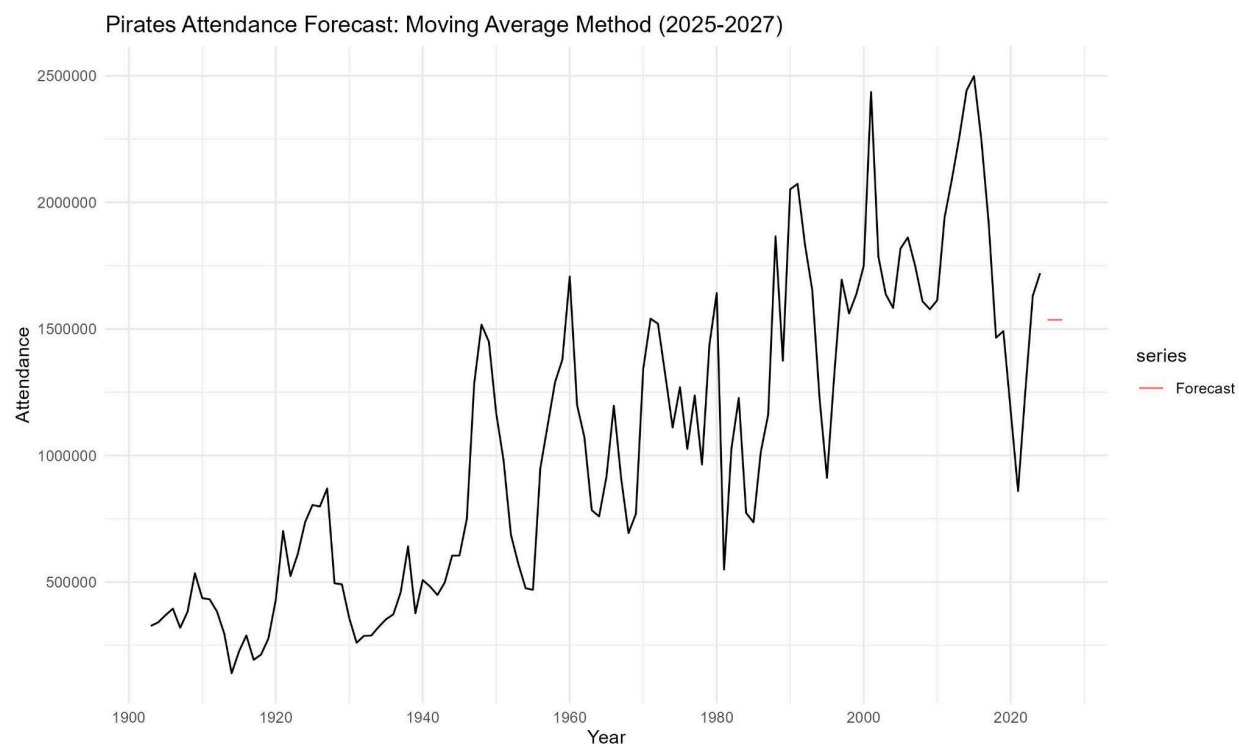
References

"Baseball History, American History and You." *Baseball History, American History and You | Baseball Hall of Fame*, National Baseball Hall of Fame, baseballhall.org/baseball-history-american-history-and-you. Accessed 27 Feb. 2025.

"Consumer Price Index Data from 1913 to 2025." *US Inflation Calculator | Easily Calculate How the Buying Power of the U.S. Dollar Has Changed from 1913 to 2023. Get Inflation Rates and U.S. Inflation News.*, 12 Feb. 2025, www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/.

"2025 MLB Team Salary Payroll Tracker." *Spotrac.Com*, Spotrac, 2025, www.spotrac.com/mlb/payroll/_/year/2025/sort/cap_total2.