The Effect of Travel on Baseball Performance

Abeil Fattore
University of Colorado, Boulder
abiel.fattore@colorado.edu

Brandon Zink
University of Colorado, Boulder
brandon.zink@colorado.edu

Cameron Connor
University of Colorado, Boulder
cameron.connor@colorado.edu

Abstract

There are 30 teams in the MLB, spread out across the nation, playing 162 games per season. 81 of these games are away, split into about 30 different series (2-3 game stints). This means a lot of travel for baseball players in order to get to all of the games, with very little down time.

We aim to look at the effect of all of this travel on a players ability to perform, both in hitting and pitching, in a variety of different common statistical categories.

Our findings indicated that there are certain distances in miles and time zones that shows a more significant effect on baseball performance (such as traveling 3 time zones west before a series). Our findings also showed that certain statistics were more greatly impacted by travel than others (such as hits per 9 innings by pitchers).

Introduction

We took our original question of 'Does travel have an effect on baseball performance', and split it into four sub-categories. We first split by batting and pitching to see if travel had a greater impact on the pitchers or batters of away teams. We then split our question into travel in terms of mileage, with groupings at the <500 miles, 500 - 1,000 miles, 1,000 - 2,000 miles, and >2,000 miles categories. Lastly, we also split our question by time zones traveled, with a distinction between traveling east and traveling west, resulting in groupings from 3 hours west to 3 hours east, with a category at each time zone.

This gave us a clearer view of what the exact cause of the difference is, whether it is jet-lag caused by time zone change or time spent on a plane by distance traveled in miles.

We aimed to answer these question because they have the potential to put certain teams at a distinct competitive advantage or disadvantage. In a league where total revenue is approaching $10 billion [1], a few games won or lost because of travel can make a significant change in overall revenue, as well as the ability to make postseason play.

By identifying factors in travel that can cause

---

[1] "MLB League Revenue 2001-2017 | Statistic." *Statista*, www.statista.com/statistics/193466/total-league-revenue-of-the-mlb-since-2005/.

a change in overall performance of players, we can help indicate to teams how to change travel methods in order to limit the most performance costly trips. This can also help indicate to the MLB changes in scheduling that could help to even the playing field in terms of travel (such as forcing east coast teams to take more west coast trips).

By using our statistic factors at a given distance, we can directly estimate the amount of wins a team has or will give up over an entire season. This shows that travel can have a direct impact on who makes the playoffs based on a few games won or lost because of travel.

This data also could be used in sports betting. Over an entire season, a few extra dollars could be won by using our data to predict the change in expected runs scored, and therefore, overall wins. This would take a large sample size of bets, but does have the potential to make an impact.

Related Work

Previously, there has been one study that focused on the effect of traveling on a team's statistics, "How jet lag impairs Major League baseball performance" by Ravi Allada [2]. The study looked at 46,535 MLB games that were played between the years of 1992 and 2011. The study looked at both offensive and defensive statistics on team winning percentages and individual player statistics.

Comparing the home and away performance, the study had concluded that MLB teams that had athletes traveling eastward had a decline in performance while teams that had athletes traveling westward had no significant change.

Basically, Allada is saying that jet lag had a greater negative impact when traveling from the west to the east. Extending from that, the decline in performance when traveling eastward had to do more with defensive performance. This is similar to what we are trying to figure out except we are not checking to see if the direction of travel has an effect on baseball statistics but just if traveling in general has an effect.

Data Set

Our original data set was Retrosheets data, using primarily the events data from http://www.retrosheet.org/. This data set contains the play by play data from every single at bat in the MLB since 2010. This included the batter, pitch count, pitch sequence, pitcher, what player was in which defensive position, if there was defensive, hit type, and much more.

There were two data types, symmetric binary and numeric. Symmetric binary consisted of statistics like away or home, hit or out, etc... Numeric data included statistics like hits, earned runs, and runs scored. In order to get the data that we needed for our analysis, we used an SQL script to pull out the pertinent data and grouped them into game series. This included, but was not limited to, home and away teams, runs scored, hits, walks, strikeouts, and types of hits. We eliminated any of the data that had missing sections, plays that involved a fielding error or sacrifice hit, or games that had rainouts.

---

[2] Rachael Lallensack Jan. 23, 2017 , 5:00 PM et al. 2017. Jet lag puts baseball players off their game. (December 2017). Retrieved March 5, 2018 from http://www.sciencemag.org/news/2017/01/jet-lag-puts-baseball-players-their-game

This resulted in a .csv file that had over 65,000 objects, each with over 40 attributes. We ran our main analysis on this data set. An excerpt of the event by event data is also included in our repository for example purposes, but the entire file is much larger and is housed in a MySQL database.

We also used stadium information in order to calculate the travel distance and time zones per game. The time zone for each data was retrieved manually from http://www.mlb.com/. The latitude and longitude (used for calculating the distance in miles) was retrieved from http://thirdlanding.com/longitude-latitudes-mlb-stadiums/. Our stadiums data also included stadium codes as used from the Retrosheets dataset for using in matching games to stadium.

We chose not to go any earlier than 2000 in our data in order to try and eliminate changes in travel style or technology that may be involved in travel recovery, while still leaving us with a large enough sample size to perform an adequate analysis.

Main Techniques Applied

After getting our data set as described in the section above, we began by grouping the data into subcategories. This resulted in a data matrix resembling the following:

	Batting	Pitching
Distance (mi.)		
Time Zones		

Within each distance subsection, we had the appropriate categories for distances as described at the beginning of this paper, as well as the appropriate categories for time zones.

Upon having the data in the appropriate subsections, we then computed the averages for a variety of common baseball statistics, as described below.

Batting

R/9: Runs per nine innings, the amount of runs a team scores if the game were exactly nine innings long (most, but not all games are).

H/9: The number of hits a team has per nine innings, not weighted for hit type. This does not include sacrifice hits or reaching the base on a fielding error.

BB/9: The number of base-on-balls (or walks) a team has per nine innings, not including hit by pitch or intentional walk. This occurs by getting four called balls in an at bat.

K/9: The number of strikeouts a team has per nine innings. This occurs by swinging and missing on a third strike or getting a called third strike.

BA: Batting average. This is defined as $BA = \frac{Hits}{At\ Bats}$, and is one of the most common stats in baseball.

OBP: On-base percentage. This is the percent of the time a player reaches base (basically BA + walks), but is defined as $OBP = \frac{Hits + Walks + Hit\ By\ Pitch}{At\ Bats + Walks + Hit\ By\ Pitch + Sacrifice\ Flies}$.

We then wanted to see the effect this had on

SLG: Slugging percentage. A measure of a player's batting average, weighted for extra base hits. Defined as $SLG = \frac{1B+(2B*2)+(3B*3)+(HR*4)}{At\ Bats}$.

ISO: Isolated power. A measure a players raw power. Defined as $ISO = SLG - BA$.

Pitching

H/9: A measure of how many hits a teams pitching staff gives up per nine innings. Does not include reaching base on error or sacrifices.

BB/9: A measure of how many base-on-balls a teams pitching staff issues per nine innings. Does not include hit by pitch or intentional walks.

K/9: A measure of how many strikeouts a teams pitching staff has per innings.

ERA: Similar to R/9 for batting, but this only measures the earned runs per nine innings (eliminates runs allowed from defensive error). One of the most common pitching stats in baseball.

ISO: The same definition as batting ISO, but instead the ISO a pitchers gives up.

After calculating this values for every MLB series, we then compared the statistics within each distance and time zone subsection against the stats from the entire MLB using a T-test. This resulted in an average (so we could see if it was higher or lower than the MLB average) and a p-value. We used a significance value of .05 to decide which distance or time zone subsections contained stats which were significantly different.

winning percentage. By using the Pythagorean Winning Percentage Formula, defined by $W\% = \frac{Runs\ Scored^2}{Runs\ Scored^2 + Runs\ Allowed^2}$, we estimated how many runs a team 'should have' scored had they not been traveling for that game. We did this by taking the actual runs scored for each away game, and then multiplying them by the 'run ratio' at that distance and time zone. The run ratio for batting is defined by $\frac{MLB\ Runs\ Average}{Distance/Time\ Zone\ Runs\ Average}$. This gives us an estimate of the number of runs that team would've scored had they not been traveling in that game. We did the same for runs allowed, by instead multiplied those by the pitching run ratio, which is defined as $\frac{Distance/Time\ Zone\ Runs\ Averages}{MLB\ Runs\ Average}$. This allows us to estimate the amount of runs the team would've given up had they not been traveling that game.

We then compiled the adjusted runs and runs allowed by team and, using the Pythagorean Winning Percentage Formula, estimated their winning percentage. We then compared that to their actual winning percentage over that time to find the net difference in winning percentage.

This directly showed us how many games a team was winning or losing based on traveling condition alone, although only taking into account runs (as there is no good formula for winning percentage using a variety of statistics).

Key Results

The following tables contain the results from our statistical analysis on distance in miles and time zones. The first numeric column contains the MLB average for that stat. If a number is in red font, it means it is lower than the MLB average. If it is in green, it is higher. If the box is highlighted red, it means the statistic has a significant p value (<.05).

Statistic	MLB Averages	< 500 miles p val	< 500 miles avg.	500 - 1000 miles p val	500 - 1000 miles avg.	1000 - 2000 miles p val	1000 - 2000 miles avg.	> 2000 miles p val	> 2000 miles avg.
Batting R/9	4.469	0.981	4.470	0.185	4.506	0.330	4.495	0.002	4.363
Batting H/9	8.868	0.577	8.851	0.003	8.959	0.956	8.869	0.001	8.748
Batting BB/9	3.159	0.863	3.156	0.133	3.133	0.885	3.162	0.044	3.203
Bating K/9	6.904	0.802	6.909	0.056	6.858	0.009	6.965	0.140	6.861
Batting BA	0.251	0.568	0.251	0.001	0.254	0.787	0.251	0.002	0.249
Batting OBP	0.315	0.723	0.315	0.077	0.317	0.880	0.315	0.106	0.314
Batting SLG	0.395	0.459	0.396	0.025	0.398	0.980	0.395	0.000	0.389
Batting ISO	0.144	0.112	0.145	0.394	0.144	0.799	0.144	0.000	0.140
Pitching H/9	9.408	0.037	9.344	0.001	9.519	0.101	9.461	0.000	9.257
Pitching BB/9	3.530	0.097	3.499	0.154	3.502	0.190	3.555	0.016	3.588
Pitching K/9	6.696	0.103	6.733	0.001	6.618	0.202	6.726	0.898	6.699
Pitching ERA	4.261	0.655	4.272	0.221	4.293	0.335	4.286	0.000	4.146
Pitching ISO	4.582	0.214	4.603	0.348	4.565	0.873	4.579	0.751	4.575

Statistic	MLB Averages	3 time zones west p val	3 time zones west avg.	2 time zones west p val	2 time zones west avg.	1 time zone west p val	1 time zone west avg.	0 time zones p val	0 time zones avg.	1 time zone east p val	1 time zone east avg.	2 time zones east p val	2 time zones east avg.	3 time zones east p val	3 time zones east avg.
Batting R/9	4.469	0.000	4.195	0.711	4.488	0.001	4.591	0.499	4.453	0.013	4.558	0.074	4.558	0.079	4.549
Batting H/9	8.868	0.000	8.525	0.373	8.818	0.000	9.030	0.141	8.828	0.000	9.026	0.506	8.904	0.265	8.924
Batting BB/9	3.159	0.792	3.167	0.432	3.134	0.955	3.160	0.212	3.179	0.002	3.090	0.285	3.193	0.000	3.271
Bating K/9	6.904	0.000	7.073	0.725	6.888	0.343	6.874	0.001	6.973	0.021	6.832	0.092	6.831	0.006	6.795
Batting BA	0.251	0.000	0.244	0.298	0.250	0.000	0.255	0.090	0.250	0.000	0.255	0.721	0.252	0.229	0.253
Batting OBP	0.315	0.000	0.309	0.366	0.314	0.000	0.319	0.519	0.315	0.087	0.317	0.679	0.316	0.006	0.319
Batting SLG	0.395	0.000	0.381	0.255	0.392	0.000	0.404	0.799	0.395	0.003	0.400	0.467	0.397	0.438	0.397
Batting ISO	0.144	0.000	0.137	0.357	0.142	0.000	0.149	0.338	0.144	0.217	0.145	0.403	0.145	0.816	0.144
Pitching H/9	9.408	0.000	9.105	0.029	9.538	0.000	9.598	0.007	9.331	0.000	9.628	0.417	9.456	0.092	9.317
Pitching BB/9	3.530	0.475	3.507	0.190	3.577	0.175	3.496	0.441	3.517	0.468	3.548	0.472	3.505	0.000	3.651
Pitching K/9	6.696	0.648	6.677	0.001	6.554	0.000	6.512	0.000	6.809	0.029	6.628	0.331	6.653	0.000	6.869
Pitching ERA	4.261	0.000	3.994	0.736	4.277	0.000	4.391	0.685	4.252	0.015	4.344	0.099	4.339	0.119	4.329
Pitching ISO	4.582	0.000	4.444	0.535	4.602	0.960	4.581	0.376	4.568	0.000	4.718	0.827	4.589	0.243	4.616

There are a some key results that we can take from these tables.

Traveling more than 2,000 miles causes nearly every category to be significantly different. All significant batting stats get worse, however the significantly different pitching stats actually get better (lower in the categories where you want a lower number, and vice versa).

The most commonly significant statistics from distance traveled in miles are batting H/9, BA, SLG, and pitching H/9 (with at least two significant categories). However, in each of these significant categories, the averages are both higher and lower depending on the distance traveled.

Number of time zones crossed seems to have a more widespread effect depending on the number of zones traveled.

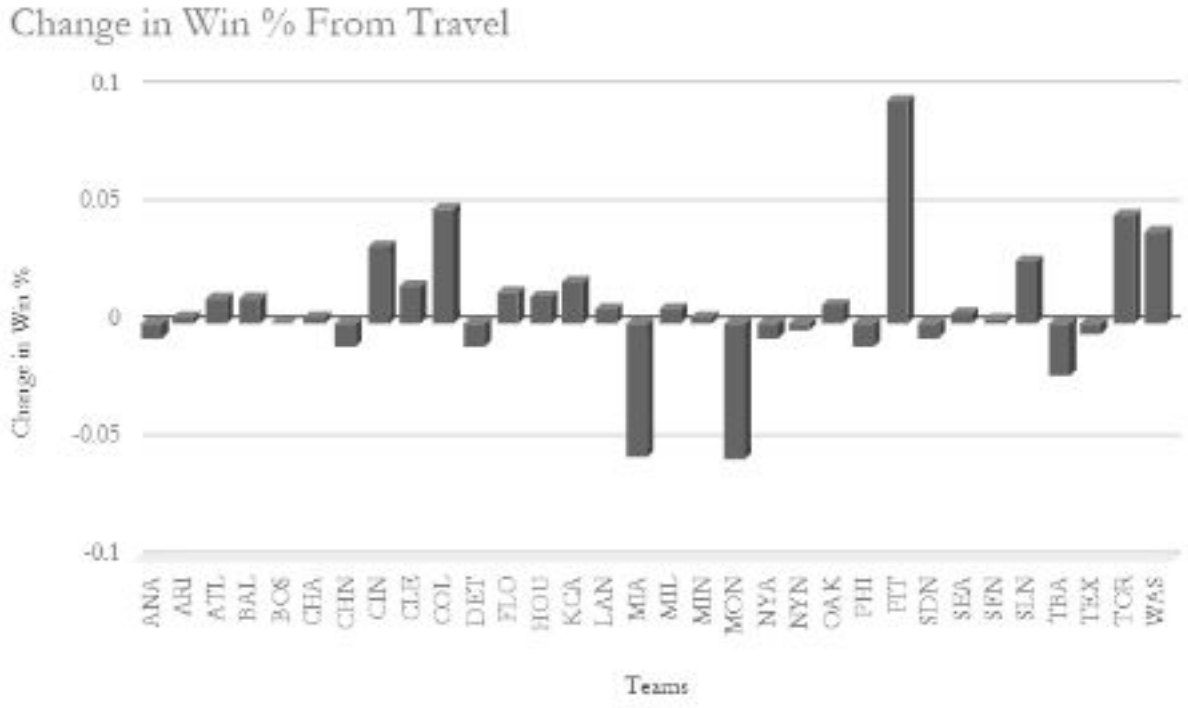
Traveling 3 hours west (from east coast to west coast) is significant in all but three categories. However, batting stats get worse, pitching stats get better. This is most likely explained by the fact that west coast games are often in the evening, meaning that east coast teams might be playing until 1 or 2 am east coast time. This can also be explained by the fact that many of the west coast stadiums (Seattle, San Francisco, Oakland, Los Angeles, San Diego) are at or near sea level, which makes it harder for the ball to travel.

Traveling 1 hours west is also significant in a number of categories, however in the inverse of 3 hours. Batting stats tend to get better, while pitching stats tend to get worse. This is most likely explained by the fact that the only area where 1 hours west games are played with frequency is from each cost to central time zone (there is only 1 mountain time zone team, making central to mountain or mountain to west rare), where many of the central parks are hitters parks (parks in which the dimensions of the outfield or the elevation generally result in higher offensive stats than most MLB parks) by happenstance.

1 hour east also had a fair number of significant categories, similar to that of traveling 1 hour west. Many east coast parks are also hitters parks, resulting in the same reasoning as from 1 hour west.

Only a few stats were widely significant in time zones analysis, being batting K/9, pitching H/9, and pitching K/9. The reasoning for this is unclear and possibly provides the most new insight and so how traveling across time zones effects players.

On the next page, there is the graph that shows the change in winning percentage from the Pythagorean Winning Percentage Formula as calculated using adjusted runs.



From this graph, we can see a few interesting things. There are two teams, the Miami Marlins and Montreal Expos (no longer an active team) that suffered about a -5% change in winning percentage per season. Over 81 away games per season, this means a change in about 4 games lost that should've been won had they not been traveling extra.

On the other side of things, the Pittsburgh Pirates saw an +10% change in winning percentage due to their lack of having to travel. With 81 away games, this averages out to about 8 games a season won where they should've lost had they been on an MLB average travel schedule. The Cincinnati Reds, Colorado Rockies, Toronto Blue Jays, and Washington Nationals all saw an increase of between 2 and 4 wins due to lesser travel.

Applications

If we look to last seasons (the 2017 MLB seasons) playoff standings at the end of the

regular season [3], we see the following standings for wild card eligible teams (there are 2 wild card teams per league.

---

[3] "Wild Card Standings." *Major League Baseball*, mlb.mlb.com/mlb/standings/wildcard.jsp.

American League

Team	W	L	GB
New York Yankees	91	71	+6
Minnesota Twins	85	77	-
Kansas City Royals	80	82	5
Los Angeles Angels	80	82	5
Tampa Bay Rays	80	82	5

National League

Team	W	L	GB
Arizona Diamondbacks	93	69	+6
Colorado Rockies	87	75	-
Milwaukee Brewers	86	76	1
St. Louis Cardinals	83	79	4
Miami Marlins	77	85	10

Just in last year alone, both wildcards were decided by margins that could have been influenced and potentially changed based on travel alone. Colorado gains an extra 4 games a year from travel factors. Milwaukee gains about 0.5 games. This means that, theoretically, Milwaukee should've had the wildcard spot had travel not been a factor.

Just making the playoffs has huge financial implications for a team and the players on that team.

Under the current MLB bargaining agreement, the players get a percentage of the revenue from the first 4 games of every series (to eliminate throwing games to extend the series). Just making the playoffs and losing in the first round nets the players 3% of 60% of the revenue that is split among the players. Winning the one wild card game and losing in the next round nets players 13% of the 60%.

The other 40% from required game ticket revenue is split evenly among every club that makes the postseason.

The highest overall earning year for the playoffs was $22 million, which means that had the Brewers made the playoffs instead of the Rockies, and lost in the wildcard round, each player would've gotten an additional ($22 million * 60% * 3% / 40 man roster) = $9,900. If they win one game, that number jumps to $42,900.

As for the club, had the Brewers made the playoffs, the club would've gotten $733,333.

This doesn't take into account going past 4 games, in which the two playing teams split the ticket revenue between the two.

There's a huge amount of money up for grabs in the playoffs, and a few games won or lost throughout the season because of travel can directly impact the overall earnings of ball clubs and players.