Home Team Scoring Advantage in the First Inning Largely Due to Time

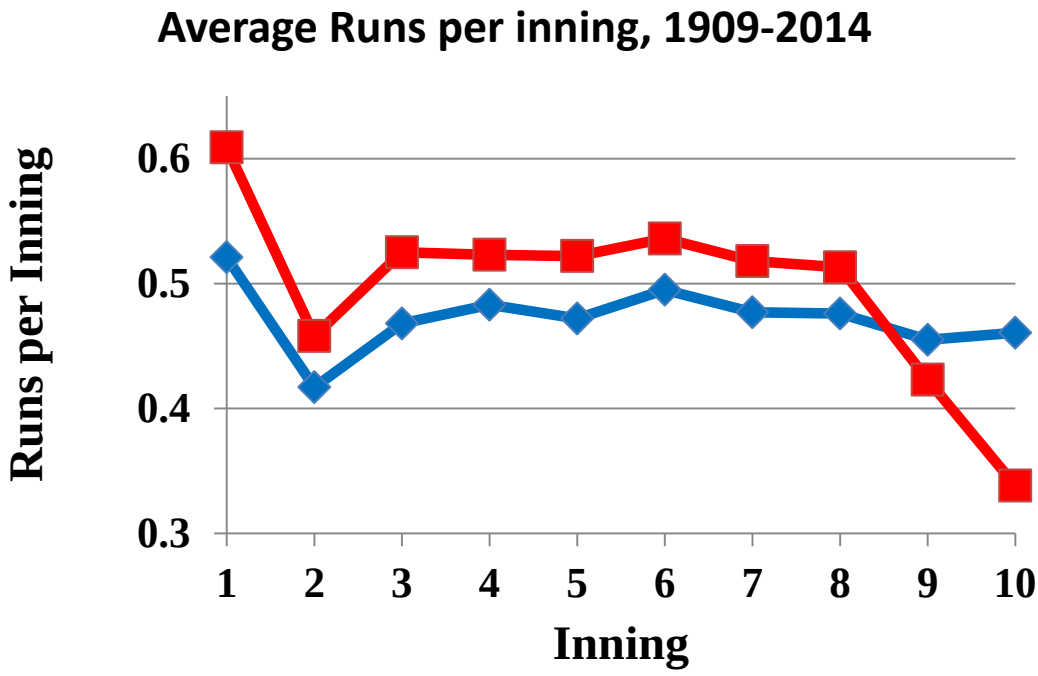
By David W. Smith
Presented June 26, 2015
SABR45, Chicago, Illinois

Throughout baseball history, the home team has scored significantly more runs in the bottom of the first inning than either team does in any other inning. At last year's SABR convention, I explored this phenomenon and found some partial explanations relating to the amount of time it takes to play the top of the first inning. That paper can be found on the Retrosheet web site at:

(http://www.retrosheet.org/Research/SmithD/WhyDoHomeTeamsScoreSoMuchInTheFirstInning.pdf).

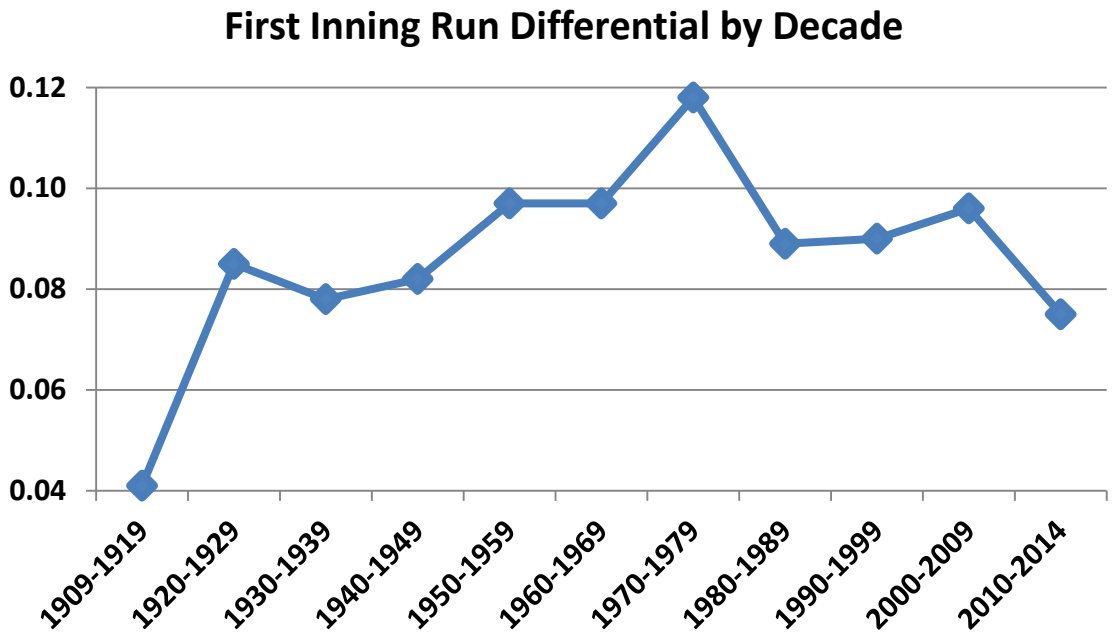
I offer a brief summary of my major findings from last year so that we have a good context for the new information I will present today. Figure 1 presents the primary observation of the pattern of run-scoring. On average, the home team scores 0.1 more runs than the visiting team, a difference which is highly significant by standard statistical tests such as t-test and Analysis of Variance. This difference accounts for 58% of the difference in scoring between the home and road teams, which is a startling result. As I noted before, home-field advantage does not derive from the fact that the home team bats last, but rather from the fact that they play the first inning!

Figure 1. (Home team data in red, visiting team in blue)



The magnitude of this effect has varied greatly over the past 107 seasons. These changes are summarized by decade in Figure 2.

Figure 2. First Inning Run Differential



There is a relationship to overall scoring, but that is not as clear as might be expected. When scoring was very low in the deadball era, the differential was at its lowest level. The introduction of the lively ball in 1920 had an obvious impact, but the only big chance since then is the large difference seen for games played in 1970s. Since this was definitely not an era of high scoring, it is unclear how this pattern arose. The last two decades certainly saw big changes in overall offense, but the effects on the first inning differential are much smaller. Clearly something besides scoring level is at work here.

The first inning is special in many ways, one of which is the leadoff batter in the lineup always bats first. I compared home team scoring in the first inning to later innings in which the first man in the batting order led off the inning. It was clear that this is a first inning effect and not a leadoff batter phenomenon.

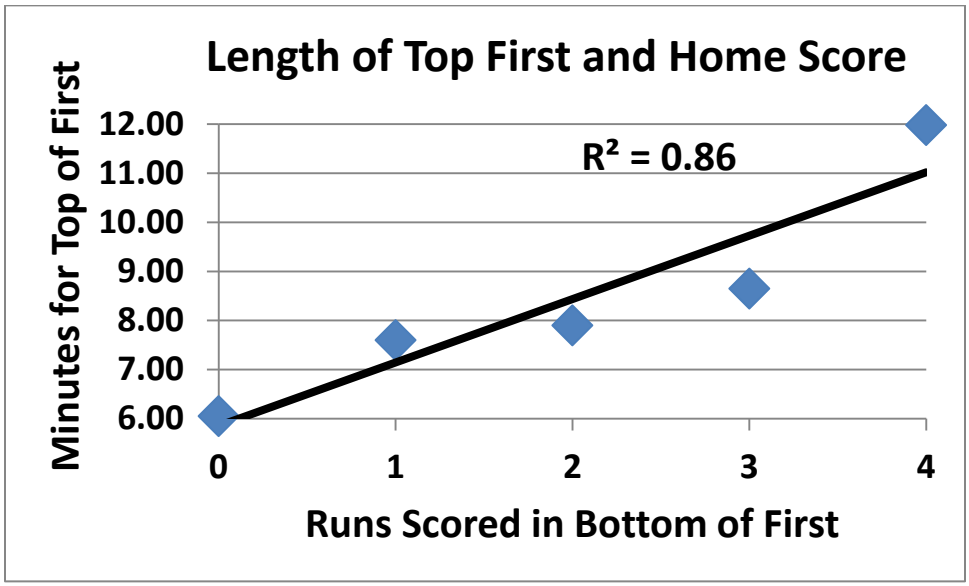
The hypothesis I developed was that the visiting starting pitcher is disadvantaged by the highly variable amount of time it can take to play the top of the first whereas the home starter knows exactly when he will take the mound. I checked three kinds of data related to the top of the first inning:

Runs scored by the visiting team
Number of batters sent to the plate by the visiting team
Number of pitches thrown to the visiting team.

All of these showed a correlation with the length of the top of the first and therefore supported my hypothesis, but of course they are only indirect measures of time. During the last year, I have discovered that that the actual time of each inning is available for recent seasons on the MLB website. They present a huge amount of data about each pitch, including pitch type and speed,

but they also record the exact time each pitch was thrown, down to the second! The most robust form of this information has only been available since 2010, but that is 12,150 games for the 2010-2014 seasons. This direct knowledge of the duration of the top of the first is exactly the data I needed to give the most rigorous test of my time hypothesis. The results are shown in Figure 3.

Figure 3. Runs scored as function of length of top of first

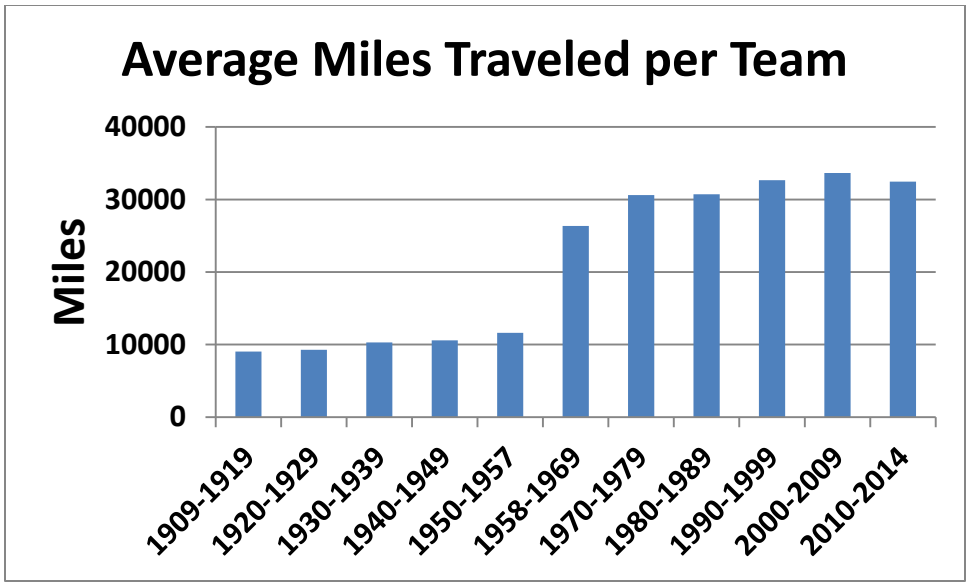


The results are compelling. The average top of the first inning with no runs (which is by far the majority of cases) takes around six minutes to play. For innings with 4 or more runs, the average elapsed time is just under 12 minutes. This agrees nicely with the indirect measures I presented last year and lends strong support to the hypothesis that the visiting starter is affected by sitting around longer before he gets his turn to pitch. The r-squared value of 0.86 means that 86% of the variance in this relationship is explained by the trend line shown. That is a very high value for baseball data.

I considered one other factor which goes beyond a single game and that is travel. In addition to adjustments to the new park and new mound, the visiting team may have great variation in getting to the new park to start the next series. If that were so, then the performance in the first game of a series might be compromised with respect to the later games in the series. Similarly, the home team may have adjustments to make when playing the first game of a home stand.

I started by calculating the distance traveled by each team in each season. Figure 4 shows a pattern which should not be surprising.

Figure 4. Average miles traveled by each team, 1909-2014



Obviously things changed a lot when teams started traveling to and from California in 1958. The average travel distance per team jumped from around 10,000 miles before 1958 to the current average of just over 30,000 miles. Of course, there is still variation around that average, as summarized in Table 1.

Table 1. Range of miles traveled

	1909			2014	
Low	Giants	7095	Low	Cubs	22933
Median	Reds	8927	Median	Nationals	31350
High	Browns	11290	High	Mariners	52519
Average		9370	Average		32560

In the current age, the average team plays 54 series with 13 home stands (there are small variations) and takes 35 trips of greatly varying total distance.

What are the effects of this travel on performance? I split all the trips taken into quartiles for the two eras since the totals are so different since 1958. We find not surprisingly that the actual distances on these quartiles are very different pre- and post-California. Table 2 presents these values

Pre-1958	Post-1957
0-125	0-350
126-260	351-700
261-400	701-1200
401-1100	1201-2800

The mode of travel is of note as well. Prior to 1958, the large majority of trips were by train and since then they are by air. The short trips today are by bus: Los Angeles-San Diego;

Philadelphia-New York; Chicago-Milwaukee. Therefore a 1000 mile trip in 1935 was very different from one in 2014 in terms of the time for the travel and perhaps on performance. In addition to calculating the distance for each trip, I also looked at whether or not there was a day off between series during the travel. I did this for each series for each team from 1909-2014.

I will report these results separately for the two travel eras. Table 3 has the data for 1909-1957.

Table 3. Runs scored by home team in first inning by visitor travel distance, 1909-1957

Quartile	Days off	Runs	Quartile Average
1	0	0.653	
1	1	0.621	0.641
2	0	0.608	
2	1	0.610	0.609
3	0	0.621	
3	1	0.614	0.618
4	0	0.645	
4	1	0.610	0.621
Overall average		0.622	
Visiting team runs in 1st		0.521	

These results do not indicate a significant effect of travel with the largest difference here being for the shortest trips. My preliminary results last years indicated there was a negative effect and my abstract reflects that. However, now that I have analyzed more deeply, the travel effect is very small. There is not even a big difference when the visitors travel without having a day off.

How about the more modern era? Table 4 has those results.

Table 4. Runs scored by home team in first inning by visitor travel distance, 1958-2014

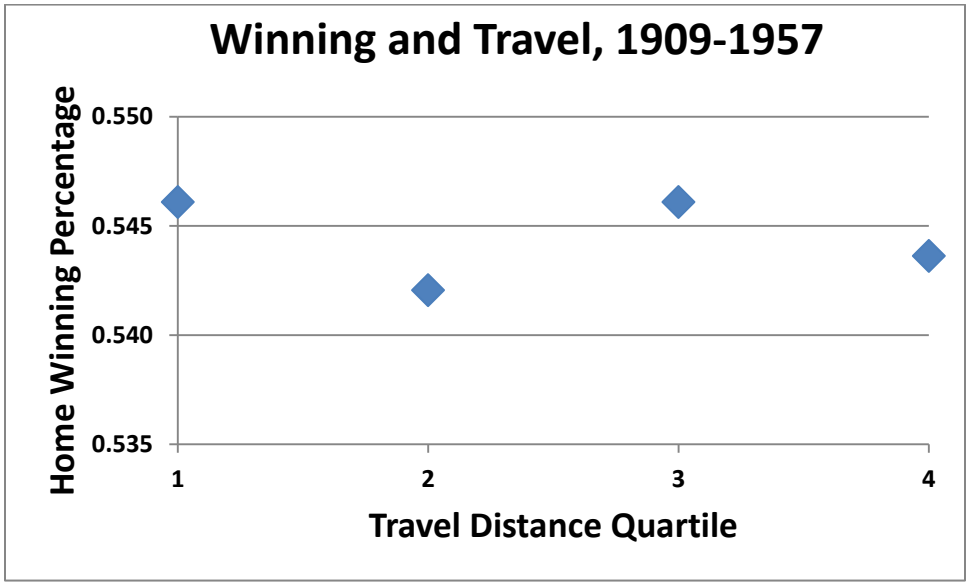
Quartile	Days off	Runs	Quartile Average
1	0	0.604	
1	1	0.602	0.604
2	0	0.612	
2	1	0.591	0.605
3	0	0.618	
3	1	0.589	0.609
4	0	0.596	
4	1	0.595	0.596
Overall average		0.603	
Visiting team runs in 1st		0.504	

These results also show no significant effect of travel, even less than for the earlier era. It did not matter which direction the long trips were (East to West or West to East). The values are

somewhat lower for the modern era (a bit less than 5%), but the absolute differential of a tenth of a run is very constant. The conclusion has to be that travel does not explain the first inning difference.

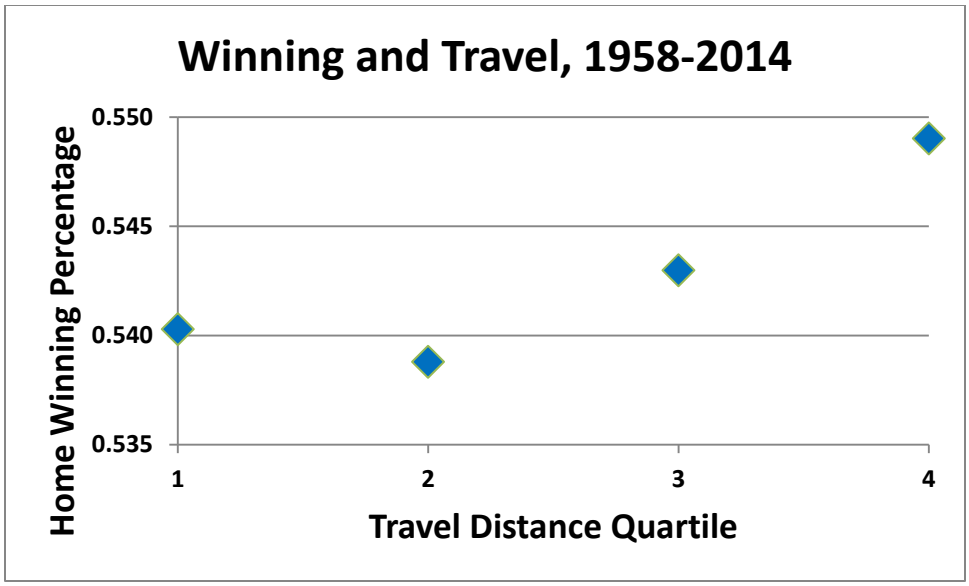
I expanded this one final way, which is to look at home team winning percentage in the first game of a series. After all, winning is what really matters and the first inning data have the biggest meaning in possible contribution to winning. Once again I present these results by travel era.

Figure 5. Home team winning percentage in first game of series as function of distance traveled by visiting team, 1909-1957.



There is no relation at all between home team winning percentage in the first game of a series and the distance traveled by the visitor in this era. I looked at having an off-day for travel and that did not matter either. How about for modern times?

Figure 6. Home team winning percentage in first game of series as function of distance traveled by visiting team, 1958-2014.



There is a small increase for the longest trips, but it is certainly not a dramatic one. The overall effect of travel on winning percentage is small at best.

Conclusions

- There is a clear and consistent home team scoring advantage in the first inning.
- Travel does not explain the scoring difference.
- Home team winning percentage also not affected by travel.
- Greater length of top of first offers the best explanation for increased scoring in bottom of first.