Why Do Home Teams Score So Much in the First Inning?

By David W. Smith
Presented August 2, 2014 at SABR44, Houston, Texas

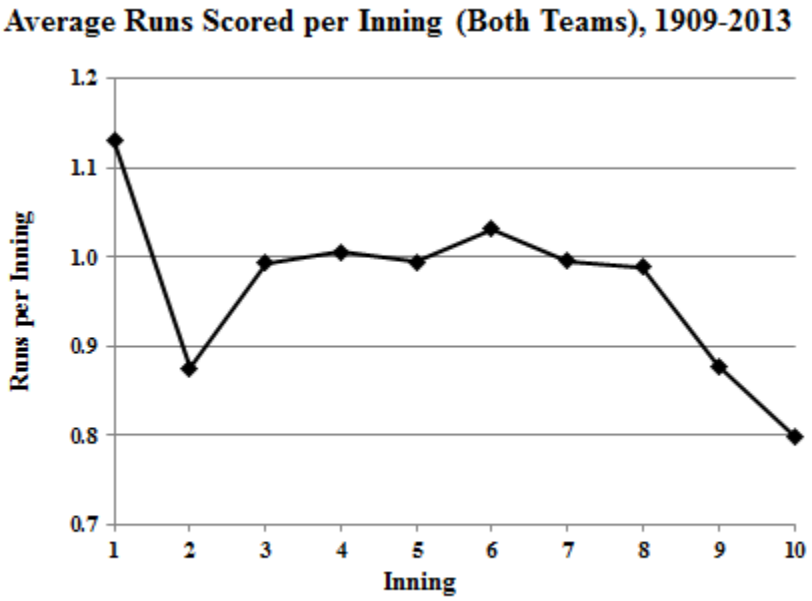
In an earlier study on batting order (http://www.retrosheet.org/Research/SmithD/Batting%20Order%20Lineup2006.pdf), I discovered striking differences in the rate at which runs are scored in each inning, with the largest effect being high values in the first inning, especially for the home team. The current paper will explore the large first inning difference in more detail with the hope of coming up with possible explanations.

All data used in this study came from Retrosheet (www.retrosheet.org) and are available free of charge on the Retrosheet web site. There are three relevant and overlapping subsets which will be used to address different questions:

Line score data	1909-2013	175,038 games
Play by play data	1949-2013	125,321 games
Pitch by pitch data	1988-2013	59,811 games

The starting point is the average number of runs scored per inning, as shown in Figure 1. This pattern is the same I found in 2006, but more seasons and more games are included in the analysis.

Figure 1.

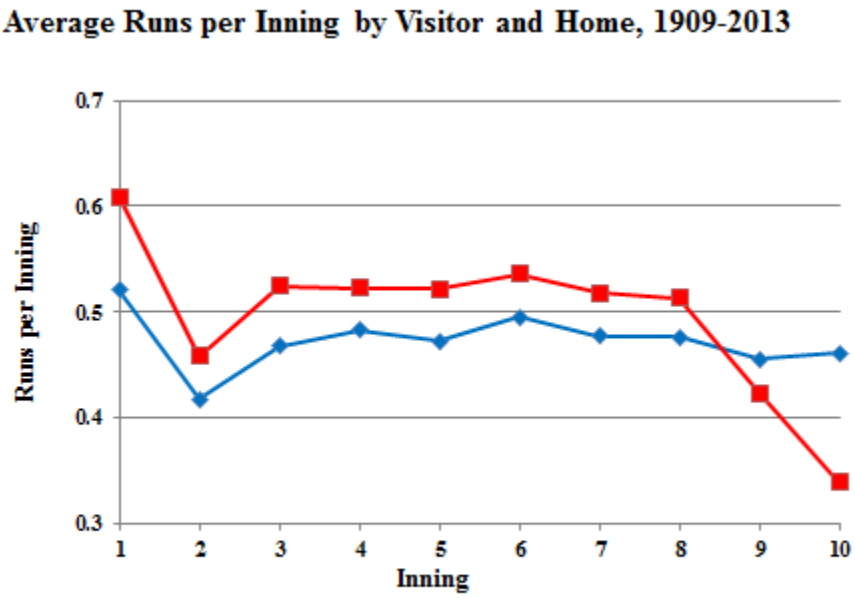


All innings beyond the 9th are combined as "10". The first and second innings stand out, although in opposite ways largely reflecting which batters come to the plate in those innings.

Aside from a small increase in the 6th inning, to which we will return, the average from innings three through 8 is remarkably constant. The decrease for the 9th inning and later reflects "walkoff" wins for the home team which of course curtail scoring since the home team stops batting in those cases, whereas the visitors have no limit on their run-scoring potential in the top half of those innings.

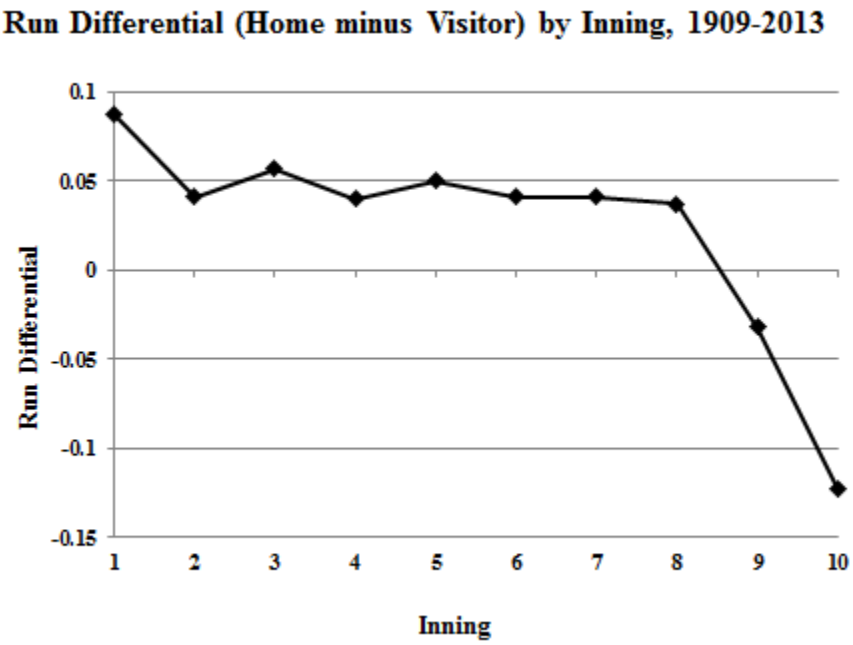
When the data are divided into visiting and home teams, we see that they score at very different rates, as shown in Figure 2 which has average runs per inning for the home team in red and for the visitors in blue.

Figure 2.



The drop off in the 9th inning and later is once again due to walkoff wins, but the lines are basically parallel before that. Another way to dissect this basic pattern is by the difference between the visiting and home teams as shown in Figure 3.

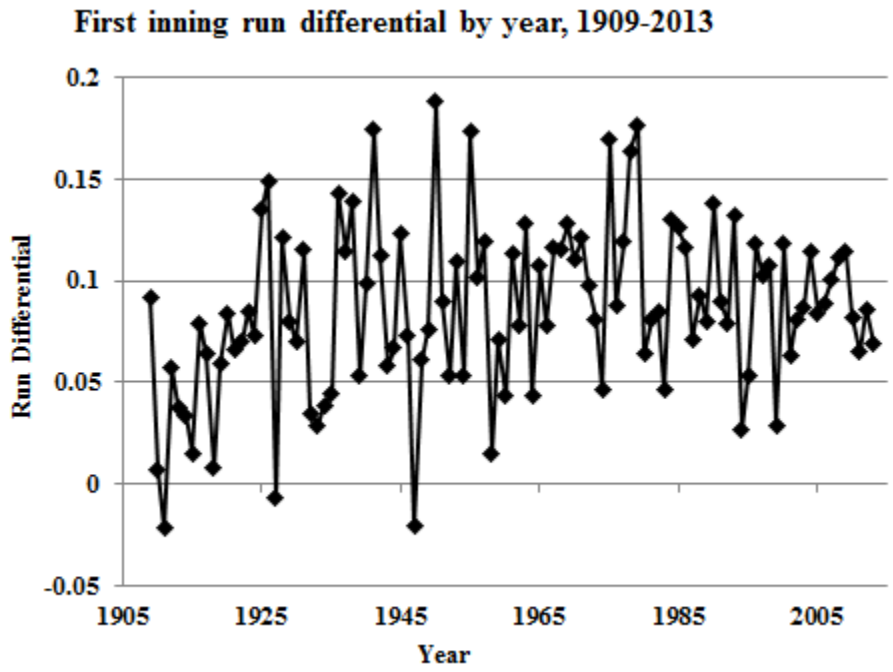
Figure 3.



The differential displayed in Figure 3 shows that the parallel appearance in Figure 2 was deceptive and the first inning stands out from the rest of the game.

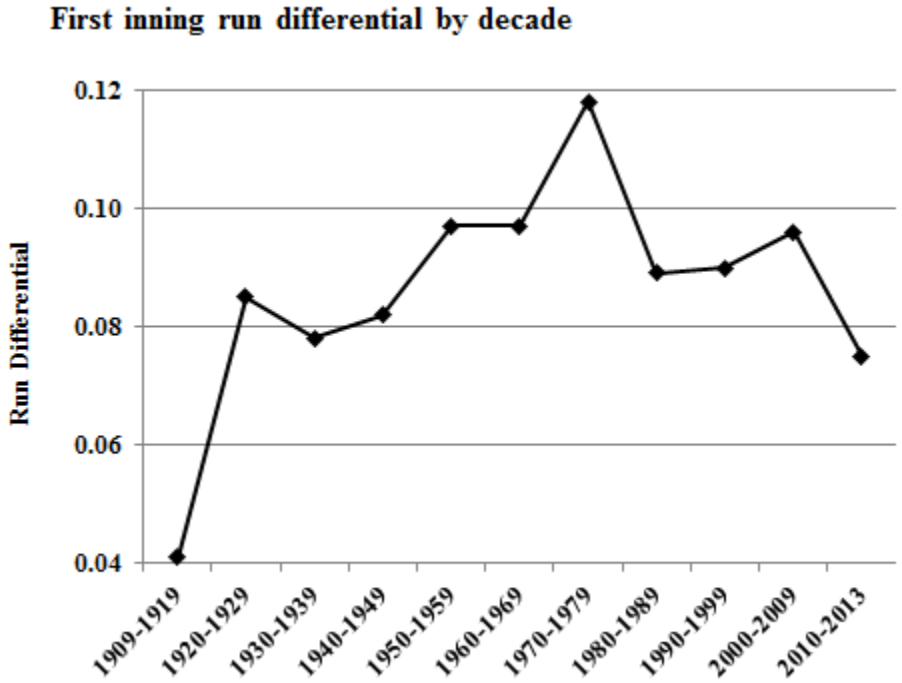
The numbers presented to this point come from all games across the last 105 seasons and I often find it valuable to analyze results in the aggregate. However, it is reasonable to ask how variable this first inning pattern has been over this long time period. Figures 4 and 5 address the question of variability. In order to make the graphs more readable, I present the differences between home and visiting teams rather than the raw data for each.

Figure 4.



As expected, there is a lot of annual variation, but in 102 of the 105 seasons the home team scores more. I decided to combine the numbers into groups by decade to see if that smoothed out the pattern and that result is in Figure 5.

Figure 5.



This consolidation decreases the noise in the data, but shows some interesting changes over the study period, from the very small difference in the dead ball era to a surprising peak in the 1970s and a clear decline in recent years. At this time, I do not have a speculation as to why these long-term patterns occur.

Of course, the most meaningful question to ask about any scoring study is the impact on winning. Table 1 addresses the question of home team advantage and the specific contribution of the first inning to this advantage, both in terms of a full season.

Table 1. Differences in total scoring, 1909-2013, normalized to 162 game season

Average total scoring, 1909-2013

Home team: 4.50 runs per game
Visiting team: 4.35 runs per game

Differential of 0.15 runs per game which is 24.5 runs over 162 game season

First inning different of 0.088 runs is leads to 14.3 runs over 162 game season

Therefore, first inning accounts for 58% of the total difference

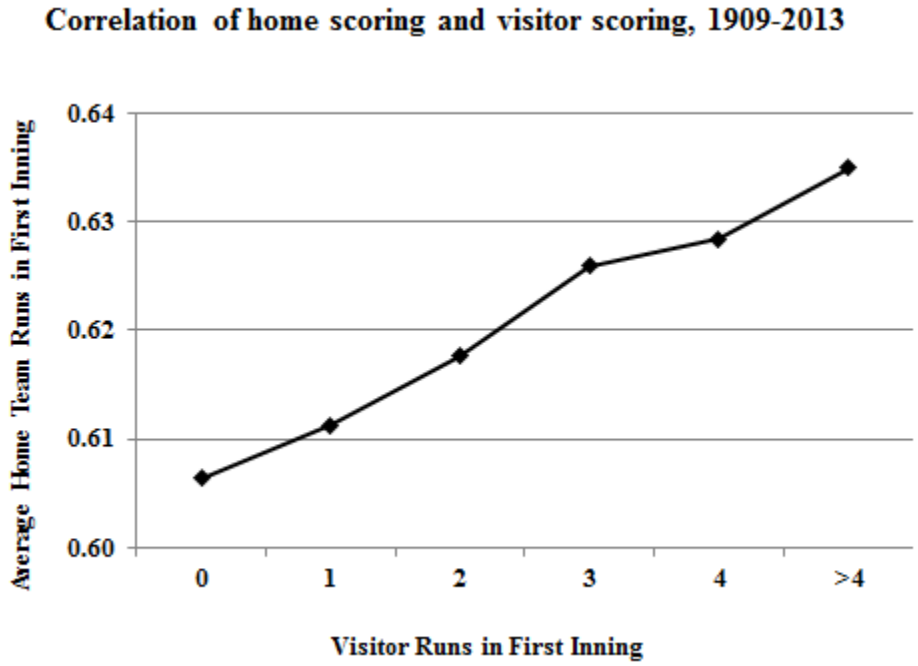
From 1909-2013, home teams won 54% of all games (94675 – 80348). Conventional wisdom has it that the home team advantage reflects the value of batting last. The results in Table 1 indicate that the real value is that they play the first inning!

What accounts for this substantial effect of the first inning? A complete answer must address not only the difference between the two sides in the first inning but also the difference between the first inning and later in the game.

The other way to look at higher offense, of course, is to see it as lower effectiveness of the opposing pitcher. It then follows from the above data that both starting pitchers have trouble to begin the game, but the visitor struggles more. My hypothesis is that the visiting starter is affected by how long he sits on the bench in the top of the first, specifically that a long rest disconnects him from his warmups and leads to problems when he does take the mound in the bottom of the first. Also, the bullpen mound could be different from the game mound.

Since we don't have actual times for the top of the first inning, I looked for other types of information that would indicate longer times for the visiting batters. The first of these surrogates for time is the relation of home team scoring in the first to the scoring in the top of the first, as shown in Figure 5.

Figure 5.



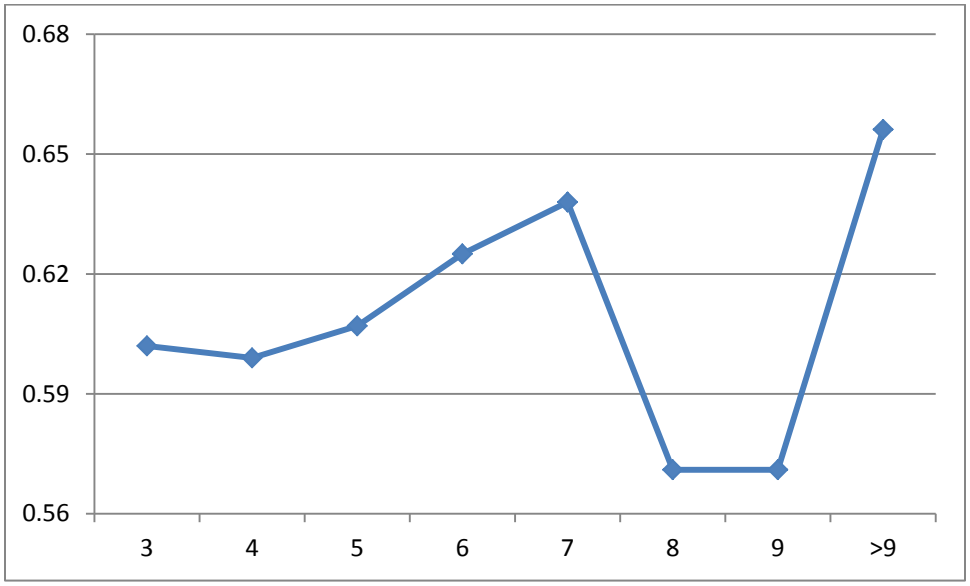
The correlation coefficient of this relationship (r) is 0.994 and r^2 is 0.988, meaning that 98.8% of the variance is accounted for by this line, which is an extraordinarily high value. It certainly takes a longer time for the visitors to score more runs, so this result supports the hypothesis, although it is perhaps counterintuitive that increased scoring by the visitors leads to a corresponding increase by the home team.

Frequency of Scoring Differences after 1st inning, 1909-2013

175,038	Games
84,660	Times both teams scored 0
38,792	Visitor led
45,691	Home led
5,895	Tied, but not at 0-0

Next I considered the relation of home scoring to the number of batters in the top of the first, as shown in Figure 6.

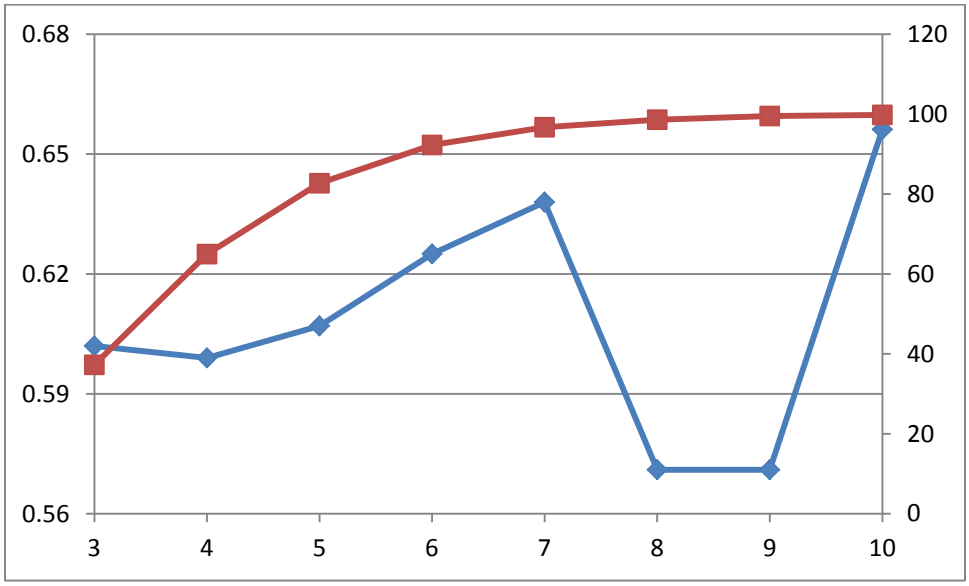
Figure 6. Average runs scored by home team in the first inning as a function of number of batters in the top of the first, 1949-2013.



Although it will generally take more time to play the top of the first when there are more batters, the relationship here is complicated. There is a very small difference between innings in which the visitors have either three or four batters, then a strong relationship between four and seven

batters. The wild swings for 8, 9, or more batters are hard to explain. I offer one explanation, namely small sample size, as shown in Figure 7, which is the same as Figure 6 with the cumulative percentage of the total on the second line.

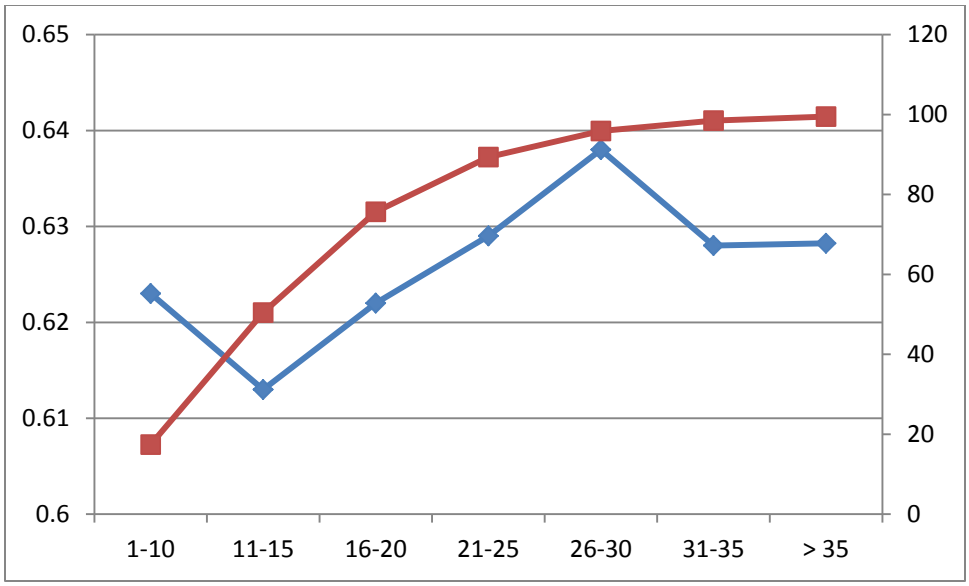
Figure 7. Average runs scored by home team in the first inning as a function of number of batters in the top of the first, 1949-2013, with cumulative percentage of appearances.



That is, first innings for the visitors with 7 or fewer batters account for 97% of the total innings and the remaining 3% show statistical noise.

Another measure of time spent in the top of the first is number of pitches thrown for which our data extends back to 1988 with the result shown in Figure 8.

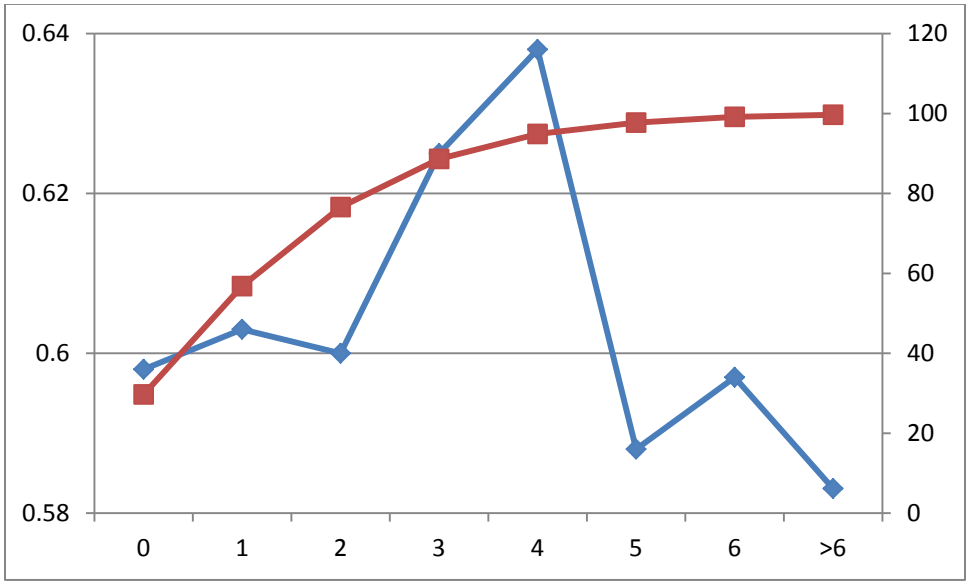
Figure 8. Average runs scored by home team in the first inning as a function of number of pitches thrown in the top of the first, 1988-2013, with cumulative percentage of each pitch range.



Since it takes more time to throw more pitches, we see once again that the general relation holds, namely that longer visitor first innings lead to more scoring by the home team in the first. There is a deviation at the high end as we saw for number of batters due to the very small number of these innings, 4% of the total. The result seen with the lowest number of pitches does not fit the expected pattern – perhaps these innings were so fast that the visiting starter did not have enough rest!

The last of these indirect measures I considered was the number of men to reach base in the top of the first, under the assumption that innings with more runners take longer to play. The result is in Figure 9.

Figure 9. Average runs scored by home team in the first inning as a function of visiting batters reaching base in the top of the first, 1949-2013.



The relation is not as strong, but is still visible, up through four men reaching base. As the cumulative percentage line shows, these cases comprise 95% of the total. Taken together, these four measures are consistent with my proposal that more spent by the visitors batting in the top of the first leads to more scoring by the home team in the bottom of the first.

In addition to changes after the first inning within a game, I also looked at changes between games, specifically comparing the first game of a series to later games. If the problems the visitors have are related to lack of familiarity with the park or the mound, then this may be exacerbated by the effect of travel and thereby be reflected in the first game of a series as well. I calculated average runs per game for the first game compared to the average for all later games for all series for all 105 seasons, 1909-2013. These results are in Table 4.

Table 4. Record in first game of series vs later games in series

Effect of First Game of Series and of Home Stand, 1909-2013

First Game of Series vs Later Games in Series

Visiting Team

In 68 Seasons, Runs per Game is lower in First Game

In 37 Seasons, Runs per Game is higher in First Game

Home Team

In 54 seasons, Runs per Game is lower in First Game

In 51 seasons, Runs per Game is higher in First Game

Home Team in First Game of Home Stand and Later Games

in 61 seasons, Runs per Game is lower in First Game

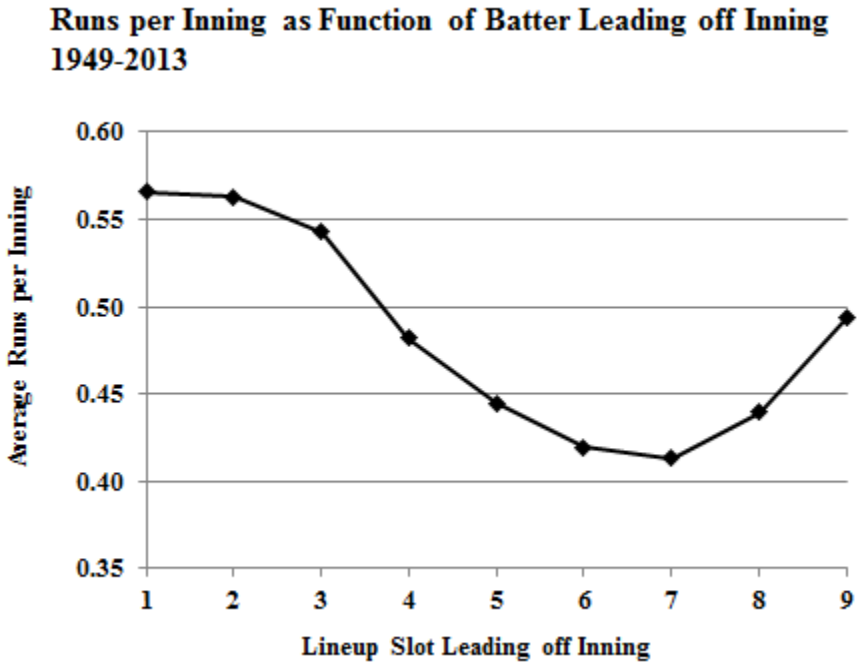
in 44 seasons, Runs per Game is higher in First Game

It appears there is an effect of traveling, in that the first game of each series followed travel by the road team, but not necessarily for the home team. The first game of each home stand definitely followed travel by the home team. Indeed, there is a travel effect, but it is a larger problem for the visitors than for the home team, which is probably not surprising.

Perhaps the most obvious feature of the first inning is which batters appear. It is the only inning in which there is a guarantee that the manager's choice of leadoff batter will actually bat first. For the present analysis, the challenge is to separate out the identity of the first inning batters from the fact that they are batting in the first inning. That is, is it the batters or the inning that matter more?

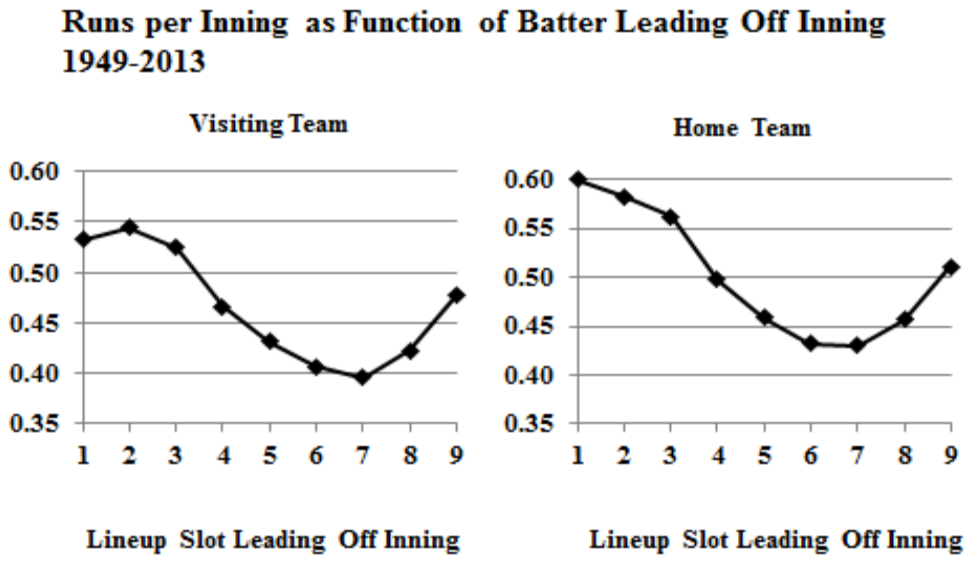
I touched on this question very lightly in my paper in 2006 (referenced at the top of this presentation) and found that scoring in a given inning is very strongly affected by which spot in the lineup bats first. Although the current study examined more seasons (65 vs 49), the basic pattern is what I saw before, as shown in Figure 10.

Figure 10.



There are many interesting features to this graph, which are discussed in my earlier paper, but for our current purposes, it is the first spot that matters the most. When the visiting and home teams are examined separately, we find some differences, as shown in Figure 11.

Figure 11.



These are remarkably different profiles with the visitors scoring fewer runs no matter who bats first. Perhaps the most surprising observation is that the visiting team does not have its maximum scoring when the leadoff man bats first.

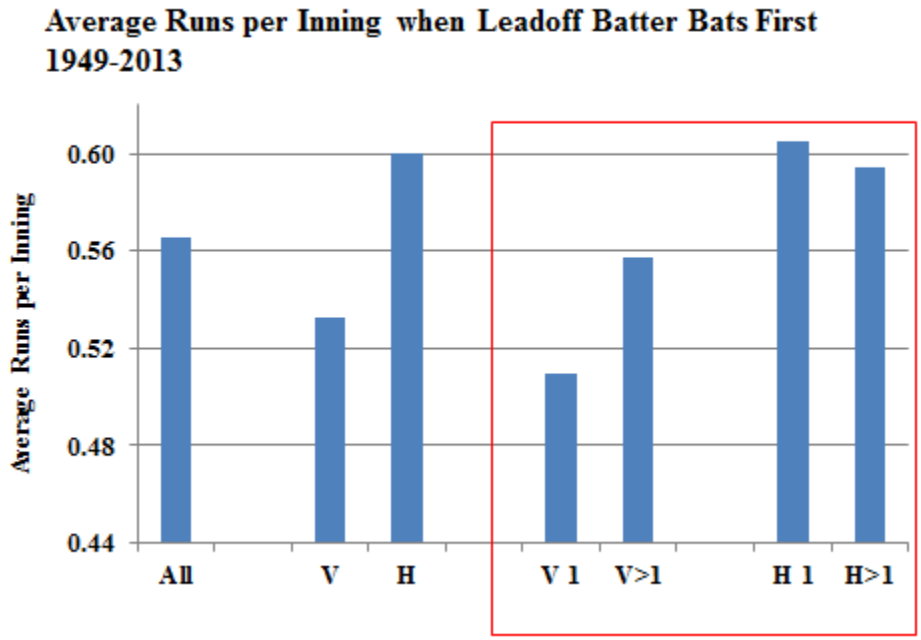
In order to tease out the effect of the leadoff batter vs the first inning, I did a series of comparisons which I will show below, but there is no reason to show graphs that have lineup slots 2 through 9 each time since they won't change. Therefore I will concentrate on the leadoff batter. First, I made a tabular comparison of the first inning performances.

Table 2. Average runs scored per inning when leadoff man bats first, 1949-2013

All Innings, Both Teams	**0.57**
All Innings, Visitors	**0.53**
All Inning, Home	**0.60**
First Inning, Visitors	**0.51**
After First Inning, Visitors	**0.56**
First Inning, Home	**0.61**
After First Inning, Home	**0.59**

These values may make more of an impact if they are presented graphically, as I do in Figure 12.

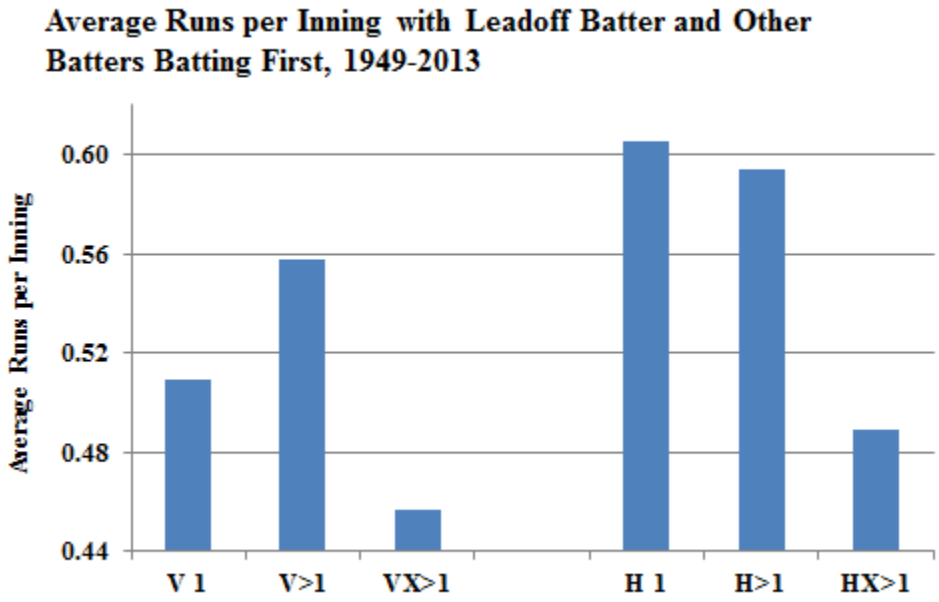
Figure 12.



I find these patterns very surprising. When the home team leadoff man bats first, his team scores at its highest per inning rate and there is very little difference between the first inning and later innings in which he bats first. However, the visiting team does much worse when its leadoff man bats first compared to later innings in which he bats first, and the first inning performance is especially low, almost 16% less than the home team. Since the visiting team still scores a lot of runs in the first inning, this pattern was unexpected.

To finish this train of thought, I redid the last calculations and added data for innings in which someone other than the leadoff batter was the first in the inning. I expanded the last two sections of Figure 12 and present the new results in Figure 13.

Figure 13.



We now have all the component pieces to sort out the effect of the first inning from the identity of the leadoff batters. Home team batters excel, or to put it in reverse terms, visiting starting pitchers do poorly. However, the visiting team batters show no such benefit from the first inning as the home starting pitchers do very well.

By the way, recall that the first graph I presented showed a small but distinct increase in scoring in the 6th inning. Another result in my 2006 paper is that the 6th inning is the only inning after the first in which the leadoff man was the most likely to bat first. This fits well with the changes during the game reported in Table 2.

In a search for a mechanistic explanation, I calculated the hit and walk rates in the first inning as compared to later innings and obtained the results presented in Table 3.

Table 3. Hits, Walks and Slugging, 1949-2013

Batter	Inning	H/Inn	BB/Inn	SA
All	**1st Only**	**1.03**	**0.42**	**.408**
All	**After 1st**	**0.98**	**0.36**	**.396**
Visitor	**1st Only**	**0.98**	**0.39**	**.391**
Visitor	**After 1st**	**0.97**	**0.35**	**.388**
Home	**1st Only**	**1.07**	**0.45**	**.424**
Home	**After 1st**	**0.99**	**0.38**	**.402**

Here we see some interesting relative changes. The overall walk and hit rates per inning decline after the first inning, implying adjustment problems for both starting pitchers at the beginning. However, the rate of change is very different for the visiting and home batters. To reverse it to the pitcher's perspective, the visiting pitchers allow 15% fewer hits per inning after the first and a WHIP that is over 9% less. The corresponding numbers for the home pitchers are 11% and about 4.5%. So on a relative basis the visiting pitchers improve more than the home team, but on an absolute level, it must be noted that the home batters started at higher rates and maintain them throughout.

Table 4. Percentage Changes in Hits, Walks and Slugging after first inning, 1949-2013

Batter	Inning	H/Inn	BB/Inn	SA
All	**1st Only**			
All	**After 1st**	**-4.47**	**-13.51**	**-3.09**
Visitor	**1st Only**			
Visitor	**After 1st**	**-1.83**	**-11.11**	**-0.76**
Home	**1st Only**			
Home	**After 1st**	**-6.83**	**-15.41**	**-5.07**

One of the strengths of this study has been the use of very large data sets, thereby allowing small differences to be detected. I am usually reluctant to subdivide the information into smaller groups, but there is one way in which I thought it would be useful in the present case. Specifically I wondered if there were some way to extend the above results to the level of individual pitchers. I therefore identified all pitchers who had at least 50 starts on the road and 50 starts at home during the 1949-2013 period and calculated their home and road first inning performances. I chose ERA as the measure although other metrics such as OPS allowed could also be used. There were 766 pitchers who met the selection criteria and their results are summarized in Table 5.

Table 5. Home and road performances in first inning for frequent starters, 1949-2013.

766 pitchers with minimum of 50 starts on road and 50 starts at home, 1949-2013

555 had higher first inning ERA on road
211 had higher first inning ERA at home

	First Inning ERA			
	Road	**Home**	**Total**	**Career ERA**
Higher on Road	**4.95**	**3.66**	**4.30**	**3.82**
Higher at Home	**4.14**	**4.82**	**4.48**	**3.99**
Total	**4.75**	**3.95**	**4.35**	**3.86**

By a ratio of over 2.6 to 1, these pitchers had more difficulty on the road than at home. Furthermore, of those with worse road records, their ERA differential was almost twice as large (1.29) as it was for those who did better at home (differential of 0.68).

From that group of 766 pitchers, I extracted a subset of the 23 who are in the Hall of Fame, including two of this year's inductees: Greg Maddux and Tom Glavine. Their data are in Table 6.

Table 6. Home and road performances in first inning for HOF pitchers, 1949-2013.

	First Inning ERA			Career ERA
	Road	**Home**	**Total**	
Higher on Road	**4.39**	**3.14**	**3.76**	**3.17**
Higher at Home	**3.54**	**3.89**	**3.71**	**3.21**
Total	**4.21**	**3.29**	**3.75**	**3.18**

Once again the clear majority (3.6 to 1) have more difficulties on the road, but the magnitude of the differences is comparable to the larger group. These superior pitchers compile an overall lower first inning ERA, but their aggregate first inning value of 3.75 is still well above their career aggregate of 3.18.

The last point on the Hall of Fame pitchers is to examine the extremes, namely those with the largest disadvantage as a visitor and those with the largest disadvantage at home. Table 7 presents the three most extreme of each type.

Table 7. HOF pitchers with greatest home-road differential in first inning ERA, 1949-2013.

	First Inning ERA			Career ERA
	Road	Home	Total	
Bob Gibson	**3.53**	**4.49**	**4.00**	**2.91**
Dennis Eckersley	**3.91**	**4.33**	**4.12**	**3.50***
Phil Niekro	**3.30**	**3.63**	**3.46**	**3.35**
Nolan Ryan	**5.38**	**3.04**	**4.14**	**3.19**
Early Wynn	**5.82**	**3.11**	**4.43**	**3.54**
Sandy Koufax	**4.78**	**1.82**	**3.35**	**2.76**

***Eckersley's career ERA as a starter was 3.71**

All of these men were tremendous starting pitchers (Eckersley is in a special category, of course and his career ERA as a starter was). The values highlighted in road are really striking. The first inning road ERA in excess of 5.00 for Wynn and Ryan is hard to believe, but the 1.82 for Koufax is astonishing as well and his differential is the largest I found for the Hall of Famers. It has been shown many times that Koufax benefitted greatly from pitching in Dodger Stadium, and this value is another aspect of that. By the way, there are two other Dodger Hall of Famers on that list, Don Drysdale and Don Sutton (who had 359 of his 583 starts for LA) and both had a definite home field advantage in first inning ERA.

Conclusions

Clearly higher scoring for home team in first inning

Much variation across years and eras

Strong correlation between home scoring and visitor scoring in first inning

Moderate correlation between home scoring and number of visiting batters, men on base in top of first and pitches in top of first

Large majority of individual starters do better at home than on road in first inning

No relation between first and later game of series

Strongest effects are elevated walks, hits and slugging in first inning

However, there is no single factor which stands out to explain the first inning difference and more analysis needs to be done.