

Consistency is key

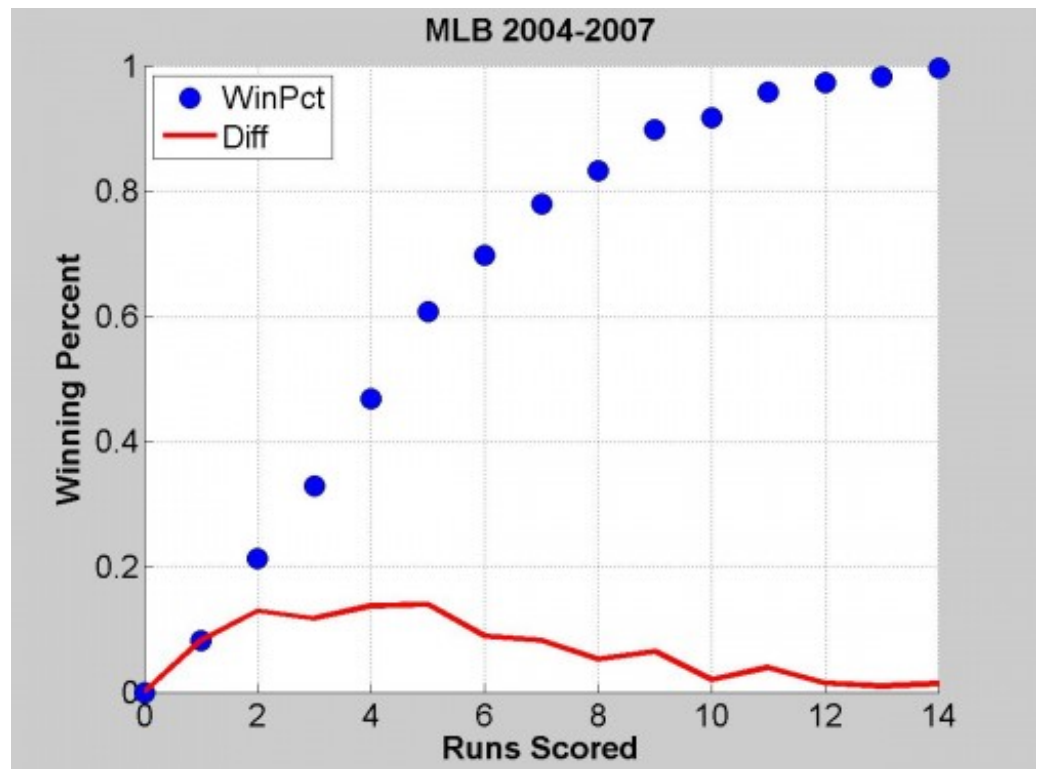
by Sal Baxamusa

December 3, 2007

You know that frustrating feeling when your team pastes the opposition by piling on 10 runs but then drops the next two games while scoring a total of three runs? Fans of every team in the majors have said it before: “Damn [insert favorite team] are feast or famine! Would it kill them to have a little consistency?”

The answer is no, it would not kill them. When it comes to scoring runs, game-to-game consistency is a good thing. Here is chart of a team’s winning percentage as a function of runs scored from 2004 to 2007:

Runs	Win Pct	Diff
0	0.000	--
1	0.083	0.083
2	0.213	0.130
3	0.330	0.118
4	0.468	0.138
5	0.608	0.140
6	0.698	0.090
7	0.780	0.083
8	0.833	0.053
9	0.898	0.065
10	0.918	0.020
11	0.958	0.040
12	0.973	0.015
13	0.983	0.010
14+	0.996	0.014



“Diff” is the change in winning percentage for at each additional run, so a team that scored five runs had a 14% greater chance of winning than a team that scored only four. Obviously, the more runs a team scored, the more likely it was to win. “Diff” peaks around five runs, which means that the marginal utility of additional runs decreased beyond this point, and at some point additional runs didn’t make much

difference at all—teams that scored 12 times basically won at the same rate as teams that scored 13 times.

With this chart, we can do a thought experiment on consistency. Imagine two teams, both who score an average of 4.5 runs in a game. The first team is full of veteran hitters who know how to play the game and do what it takes to win. In half of their games, they score four runs and in the other half they score five. Using the chart above, we can see that they would win at a .538 clip ($0.5 \times 0.468 + 0.5 \times 0.608$), the equivalent of 87 wins in a season and a possible wild card berth.

The second team is a bunch of disrespectful bonus-baby rookies who play with talent but not with heart. They score eight runs in half of their games but only one in the other half, for a .458 winning percentage. This results in 74 wins and an October full of regrets.

Advertisement

See how important it is to be consistent when scoring runs?

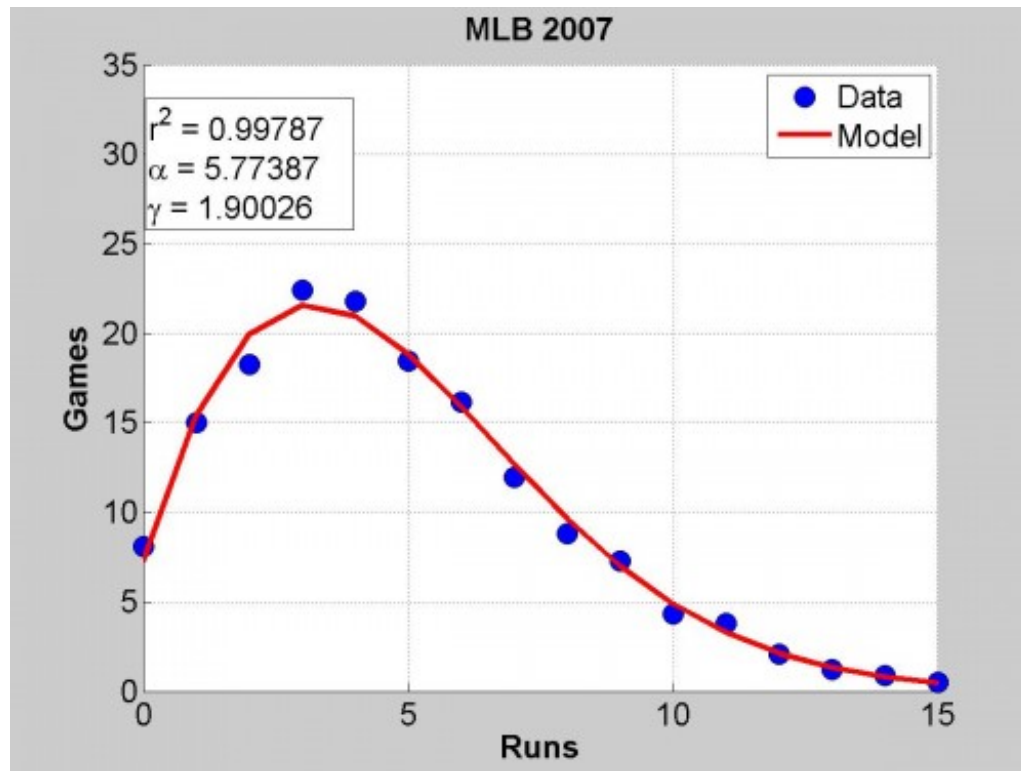
You can do the reverse thought experiment and conclude that when it comes to preventing runs, consistency is actually bad. You want a pitching staff and defense that gives up either a bunch of runs or none at all, not a host of quality-start hurling mediocrities. That is why having an ace is so important.

This is not the first time **I have mentioned this**, nor am I first person **to have noted it**. But “consistency” is one of those buzzwords that finds its way into lots of articles, commentary, and pseudo-analysis despite few actually considering its meaning. So the point bears repeating, don’t you think?

How runs are really distributed

In reality, we never see teams distribute their run scoring in quite the extreme way as in the above examples, because baseball players probably have little control over how they distribute their game-to-game offensive performances, and the collection of baseball players that we call a lineup almost certainly does not control the way they distribute their run-scoring proclivities over the course of a whole season.

This next plot shows how runs were really distributed over the course of the 2007 season.



The circles are the actual data and the line is an equation, called a **Weibull distribution**, that does a very good job of modeling run distributions in baseball. The model depends on only two parameters: the overall performance of the offense/defense and the run environment that the team plays in; the numerical values of these parameters are shown on the plot as α and γ . I'll spare you the gory mathematical details, but suffice to say that the Weibull distribution matches both actual and simulated data very well, is consistent with the Pythagorean theorem of baseball, and can account for varying run environments. You can see that the fit between the model and the data is very good; for you stat wonks, r^2 is almost one.

Advertisement

Despite the fact that there is little or no evidence that a team can control its distribution of runs scored, some teams every year do have weird run distributions. It is probably due mostly to luck. Well, not luck—let's say "unrepeatable performance that does not persist." When we call a team "inconsistent" or "feast-or-famine," we are really saying that they have a weird run distribution.

To identify which teams had the weirdest run distributions, we have to compare their actual performance against the Weibull model, which tells us what a team with the same overall performance in the same run environment would have done if not for luck...er, “unrepeatable performance that does not persist.” In other words, if a team scored three runs 25 times when the model predicted 20, then we would say they were +5 on scoring exactly three runs in a game.

The best (and worst) of 2007

With that, let’s take a look at some of the weird performances of 2007. **We did this last year as well**, but I’ve tweaked the model a little bit and I think these lists more accurately identify the “weirdness” of a team’s run distribution.

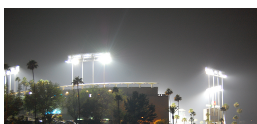
We’ll start with the teams that scored zero, one or two runs with greater frequency than expected. Based on the winning percentage by runs scored, teams scoring between zero and two runs lose a lot, so being at the top of this list is bad. I’ll present the leaders and trailers.

Scored 0–2 runs more frequently than expected

LAN	+7.4
NYA	+6.5
CLE	+4.4
...	
HOU	–5.5
SDN	–5.5
MIL	–5.5
PIT	–5.8

Well, surprise, surprise. The Dodgers, whose season was torpedoed by selfish rookies playing only for personal gain, scored between between zero and two runs in 7.4 more games than what was predicted by the model. And the Yankees, a fractured team with superstars only padding their stats for their next megacontract, were second. And then came the Indians, about whom I’m sure we can find some character flaw that explains their appearance on this list. The Padres and Brewers both missed the playoffs, but at least they went down with their guns blazing, the way only true teams do.

More from The Hardball Times



The Physics of Carry at Dodger Stadium

by David Kagan

Is the ball carrying more at Dodger Stadium?

My sarcasm is intended to show you that strange run distributions aren't really a function of team strength, tangible or otherwise. They are likely just blips—but fun blips. So, let's continue and check out the teams that poured it on, and those that didn't.

Scored 8+ runs more frequently than expected

SFN	+7.4
PHI	+6.6
MIL	+5.2
...	
TBA	-4.8
LAN	-5.5
CHA	-6.6

The White Sox had a poor offense, but at least they made their runs count by not wasting a bunch of runs in games that were likely blowouts anyway. The Giants, on the other hand, didn't make the most of the precious few runs they plated. They exploded for over seven runs in 7.4 games more than expected.

And what about our original question—which teams had consistent or inconsistent offenses? Based on the first chart, teams that scored three runs and those that scored six had winning percentages of .330 and .698. Those numbers are comparable to a **historically bad team** and a **historically good team**. Anything below three runs is truly awful, but scoring over six runs just doesn't help you that much. It's better to spread those runs out.

But we can't measure consistency based solely on how often a team scored between three and six runs. That's not fair to the very good hitting teams, who will often score over six runs, and to the very poor hitting teams, who will often score fewer than three. We have to measure it against a baseline of what we expect the team to do, for which the Weibull model is the logical choice. So, a good definition of a consistent offense is one that scores between three and six runs more often than what is expected from the Weibull model. Based on this definition, we can rank the offenses by their consistency:

Offenses, ranked by most consistent to least consistent

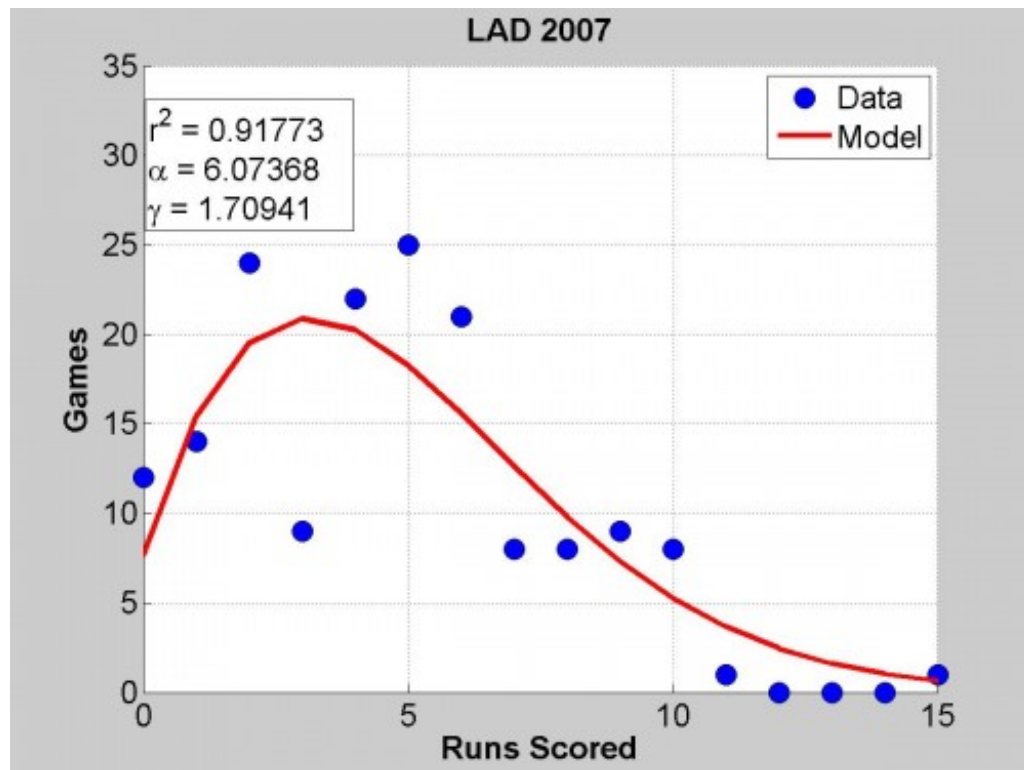
CHW	+8.6
PHI	+6.1
PIT	+5.2
KCR	+4.9
TBD	+4.8
COL	+4.3
FLO	+3.8

OAK	+3.3
ARI	+3.0
MIN	+2.9
CHC	+2.5
SDP	+2.2
LAD	+2.0
MIL	+1.9
DET	+1.5
BAL	+1.4
HOU	+1.0
SFG	+0.4
ATL	-0.5
STL	-1.0
NYM	-1.9
CLE	-1.9
CIN	-2.1
SEA	-2.4
LAA	-2.4
WAS	-2.5
TEX	-2.9
BOS	-3.8
NYN	-4.7
TOR	-5.3

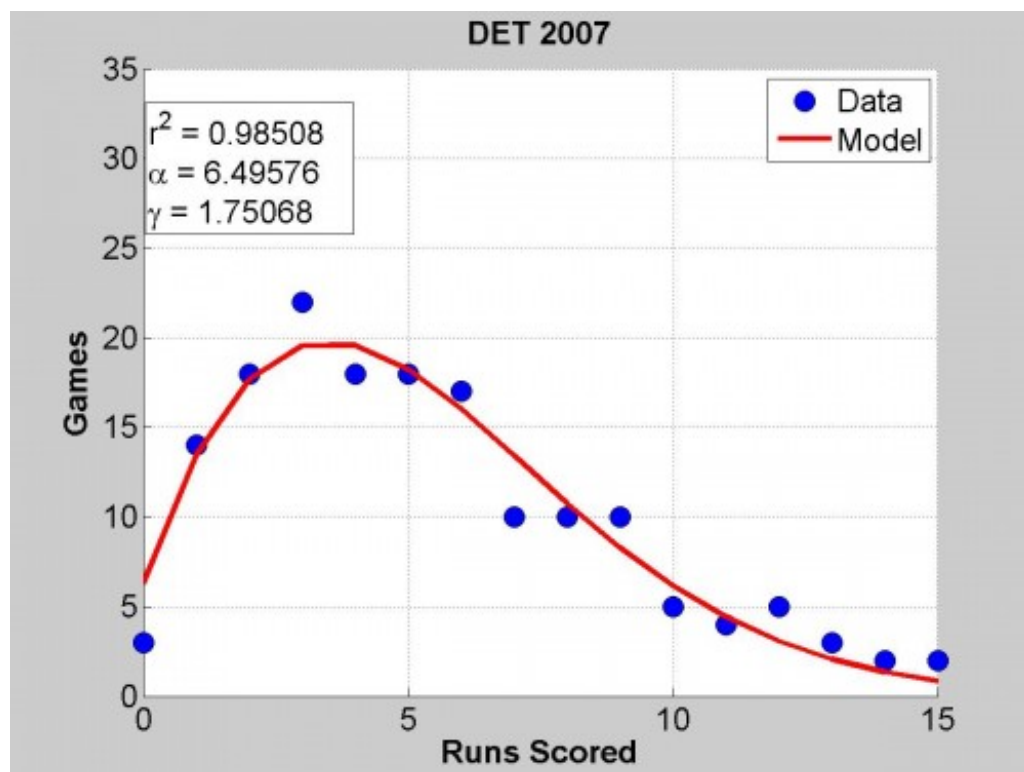
Hey, we never said that only good offenses were consistent. The White Sox and Pirates, two of the least productive offenses last year, were at least consistent. They scored in the “sweet spot”—not too many runs, not too few—in more games than expected. And an inconsistent offense is not a hallmark of a poor offense: two of the best hitting teams in baseball, the Red Sox and Yankees, featured highly inconsistent offenses. In case you are wondering, there is no statistical correlation between consistency, as measured here, and overall offensive talent, as measured by runs scored per game ($r^2=0.0001$).

Before you complain about your team’s inconsistent offense, check this list and make sure that the maddening defeats aren’t weighing heavily on your mind!

Finally, which team had the strangest run distribution of offensive performance? It was the Dodgers, who were all over the map:



On the other hand, the Tigers obeyed the model almost to a tee.



If you are interested in how your team did compared to the model, you can download plots, reports, and data from 2004-2007 by **right-clicking here and selecting “save as.”** Be forewarned: it’s about 15 MB.

Next time, we’ll take a look at the distribution of runs allowed, and we’ll dig a little deeper to see if teams do have some control over their run distributions.

I've changed the model from years past, and I don't recommend using **the old plots and reports that I made available last year**. The new model still uses the three-parameter Weibull distribution, as described **here**, but computes a separate run environment for each team's offense and defense. The shape parameter γ , which is equivalent to the Pythagorean exponent, and the scale parameter α , which is related to the overall run-scoring (or run-preventing) performance of a team, are computed by a nonlinear least-squares minimization algorithm.

My old model simply used the **Smith-Patriot model** to estimate the shape parameter and defined the scale parameter such that the model exactly matched the team's average runs scored (or allowed) per game. The old model had a bias that resulted in a non-random pattern of residuals and a systematic underestimation of the number of shutouts.

References & Resources

Something I've never done – but ought to do – is to compare the Weibull Distribution to other distributions, such as the **Tango Distribution** (scroll to bottom). It is worth reading about, though. Keith Woolner also did some work on run distributions many moons ago, **which you can check out here**.



Print This Post

Comments are closed.

The Hardball Times

Baseball. Insight. Daily.

2:14 AM ET

MembershipGamesBlogsProjectionsScoresStandingsLeadersTeamsGlossary Sign

Player & Blog Search

All major league baseball data including pitch type, velocity, batted ball location, and play-by-play data provided by Baseball Info Solutions.



FOX Sports Engage Network Partner



All minor league baseball data provided by Major League Baseball Advanced Media as distributed by STATS.

Mitchel Lichtman

All UZR (ultimate zone rating) calculations are provided courtesy of Mitchel Lichtman.

TangoTiger.com

All Win Expectancy, Leverage Index, Run Expectancy, and Fans Scouting Report data licenced from TangoTiger.com

Retrosheet.org

Play-by-play data prior to 2002 was obtained free of charge from and is copyrighted by Retrosheet.
