By the Numbers

Review

Review: "A Mathematician at the Ballpark"
Charlie Pavitt

The author reviews "A Mathematician at the Ballpark," an introduction to probability using baseball examples, which mentions a number of studies that previously appeared in BTN.

A Mathematician at the Ballpark
Odds and Probabilities for Baseball Fans

By Ken Ross

**Pi Press, 2004, 190 pages, $19.95,
ISBN 0131479903**

This is not a book about baseball. It is a book about introductory probability theory that uses some baseball examples along the way for teaching purposes. Given what it is, I thought the book was well done; Ross is described as a former math professor and I'll bet he was a good teacher. But he knows very little about statistical baseball research. For that reason, he misses good opportunities to use baseball examples; for example, in a chapter on conditional probabilities, he uses medical examples rather than probabilities for scoring given different base/out situations, which are readily available and would be an obvious choice to anyone with some relevant knowledge.

Ross seems to believe that he discovered the fact that home field advantage doesn't work during the World Series, which we've known for ages and is a central part of the data relevant to a debate in social psychological journals about whether or not there is a home field *dis*advantage in high-pressure situations.

He does refer to research with which he is familiar, such as the debate among statisticians following Albright's important article about hitting streaks. In fact, there is an Appendix in which Ross reviews some statistical baseball research articles he has happened to bump into; and I am happy to say that the bulk of the chapter consists of a short review of the chapters in our own *The Best of By The Numbers*. That led me to discover that Ross is indeed a SABR member. Ken, if you are reading this, and speaking for the Statistical Analysis Committee as a whole, thank you for the plug.

Charlie Pavitt, chazzq@udel.edu ♦

In this issue

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis – that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any – I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics: confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Shelly Appleton	slappleton@sbcglobal.net	Statistics
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
Mark E. Johnson	maejohns@yahoo.com	General
David Kaplan	dkaplan@education.wisc.edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
John Stryker	john.stryker@gmail.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@iw.net	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

How Much is a Top Prospect Worth?

Victor Wang

The value of a top prospect to the team that eventually drafts him comes from the fact that the club gets major-league performance without havnig to pay a full free-agent salary for it. How much is that benefit, in dollars? Does it vary between pitchers and hitters, or between top- tier prospects and lower-tier prospects?

Introduction

With salaries for major league free agents skyrocketing, teams are more reluctant than ever to trade their top prospects. These prospects are so valuable because if they reach their upside, a major league team has a star caliber player under their control for six full seasons while paying that player much less than what he would earn on the open market. Teams are even reluctant to trade these types of prospects for established major league stars, who may provide more certainty but cost more and may soon be free agents. I was curious to see whether teams were making the right choice by holding on to these prospects. In essence, I wanted to determine what type of value a team could get back from a top prospect during the first six years the team had that prospect under its control.

Method

To determine who the top prospects were, I examined Baseball America's Top 100 prospect lists from 1990-1999[1]. From those, I chose the top ten prospects from each year and separated them into hitters and batters. Some prospects were on the list multiple times; I only added them to my list once.

After that, I determined the WARP (wins above replacement) that they accumulated during their first six full seasons before free agency. WARP is a statistic created by Clay Davenport of Baseball Prospectus. As defined on their website, WARP is "the number of wins this player contributed, above what a replacement level hitter, fielder, and pitcher would have done, with adjustments only for within the season." While some may not agree with the baseline WARP uses, it is widely accessible for past players.

From there, I determined what the average WARP of the group of hitters and pitchers was, and broke the prospects into four subgroups. These four subgroups I called "bust," "contributor" (a back of the rotation starter or middle reliever for pitchers), "everyday player" (a middle of the rotation starter for pitchers), and "star" (an ace for pitchers). A bust was defined as a player who had 12 WARP or less (2 or less WARP per year). A contributor was defined as a player who had between 12 and 24 WARP (2 to 4 WARP per year). An everyday player was defined as a player who had between 24 and 36 WARP (4 to 6 WARP per year). A star was defined as a player who had 36 or more WARP (6 WARP or more per year). I also calculated a range for WARP by adding and subtracting one standard deviation.[2]

The results, shown in Table 1, demonstrate that teams have been getting a pretty decent return on hitting prospects. On average, the hitting prospects have given about 24 WARP, or the results of an everyday player. When that player can be controlled for a very cheap price, it gives great value to the controlling team given the current open market. However, when we take a closer look, the chances of a team getting an everyday player is one out of three. They also have a higher chance of having their prospect become a bust than getting a star player in return. A bust happens for one out of every five prospects while a team gets a star player in return for one out of every six hitting prospects. For every Vladimir Guerrero, there are even more Eric Anthonys. The large standard deviations also reflect the large risk prospects carry.

While hitting prospects provide a pretty decent return, top pitching prospects have given a terrible one. Out of the 26 different pitchers to rate as a top ten prospect, only one (Pedro Martinez) gave a star return in his first six years. A team only gets a solid starting pitcher for about one out of every ten pitching prospects. Maybe even worse, over half of the pitching prospects became busts. Given the high rate of failed pitching prospects, it could definitely be worth giving a top pitching prospect for an established player, even considering the high price that pitchers fetch in the free-agent market.

---

[1] These can be found at http://www.baseballamerica.com/today/prospects/features/26983.html .

[2] The full list of players included in this study, along with their value accumulated, is available from the author.

Considering that evaluating prospects is a subjective process, I went further down the top 100 prospects list to see if I could find similar results. This time I examined prospects rated between 11 and 25.   (Note that some of these prospects were later included in the top 10.)

These results are shown in Table 2.

Table 1 – Players Ranked as Top-Ten Prospects

Hitters

Bust	Contributor	Everyday	Star	Players	Avg WARP	WARP +/- 1 SD
10	14	16	8	48	23.72	10.96 – 36.48
21%	29%	33%	17%			

Pitchers

Bust	Contributor	Everyday	Star	Players	Avg WARP	WARP +/- 1 SD
14	8	3	1	26	12.91	0.87 – 24.95
54%	31%	12%	4%			

Comparison to Draft Studies

The study I conducted was similar to two draft studies carried out that attempted to find the average value of draft picks. These articles were written by Rany Jazayerli of Baseball Prospectus,[3] and an author who goes by the name "Philly" on the "Sons of Sam Horn" message board.[4] The latter author calculates career WARP for each draft slot, which makes it a bit difficult to compare results with my prospect study. Jazayerli also calculates career value, but he later measures the discounted value for the first 100 draft picks, which tries to estimate how much value a player gave a team before free agency.[5] Comparing his study to mine finds that a *top prospect* has a higher value than a *top draft pick*. In almost all cases, a hitting prospect rated in the top 25 would be worth more than any draftee. The only exception is with the first pick and an 11-25 hitting prospect. Hitting prospects rated in the 11-25 range have averaged about 19 WARP in the first six years which is equal to the discounted WARP of the #1 draft pick. These results would be expected, as there are more reliable scouting reports and statistics available to rate prospects. What is also interesting to see is that the top 5 picks in the draft along with a few picks after provide more value than top pitching prospects. Overall, though, top 25 prospects provide a better return than the top 25 draft picks, which is what we would expect to see.

Table 2 – Players Ranked as 11th to 25th Prospects

Hitters

Bust	Contributor	Everyday	Star	Players	Avg WARP
22	23	15	10	70	19.27
31%	33%	21%	14%		

Pitchers

Bust	Contributor	Everyday	Star	Players	Avg WARP
36	14	7	2	59	11.06
61%	24%	12%	3%		

---

[3] http://www.baseballprospectus.com/article.php?articleid=5152 . This link shows the author's main findings; there are 11 additional articles in the series.
[4] http://sonsofsamhorn.net/index.php?showtopic=4100&mode=linear
[5] http://www.baseballprospectus.com/article.php?articleid=4291

Factoring in Contracts

While we have now found the value top prospects give their teams, we have not yet factored in the lower compensation these players receive in their first six years. To see how much money these top prospects save their teams, we must determine how much value a top prospect gives to its team and for how much money. Then we must determine how much it would cost to purchase that same value in free agency. The last part is the easiest. In the *Baseball Prospectus 2006*, Baseball Prospectus determined that in the 2005 and 2006 off season, one additional WARP cost a team $1.525 million. Salary data from 1989-2007[6] shows that the average salary inflation has been 10.87%. When we factor in that inflation, on average, one additional WARP will cost a team $1.69 million in the 2007 off-season.

We have also found the value that top prospects give to their teams, so all we have to do now is determine how much it cost the teams. The new MLB labor agreement states that the minimum salary in 2007 will be $380,000, in 2008 it will be $390,000, and in 2009 it will be $400,000. The sum of these three salaries will determine how much a six year player first starting in the major leagues in 2007 will make in his first three years, assuming a team renews that player's contract each year.

The tricky part now is to find how much a player makes in years four through six. To do this I looked at every fourth, fifth, and sixth year player in the major leagues and found their salary. (All salary figures were used from Cot's Baseball Contracts.[7]) I found that the average fourth year salary was $2.13 million, fifth year salary was $3 million, and sixth year salary was $3.9 million. I then found the WARP of each fourth-sixth year player and divided their salary by their WARP. The $/WARP for a fourth year player was $.64 million/WARP, for a fifth year player it was $.83 million/WARP, and for a sixth year player it was $1.29 million/WARP.

Table 3 – Savings for Top-Ten Hitters

	Bust	Contributor	Everyday	Star
WARP (over 6 years)	6.54	18	29.63	43.35
Chance of Occuring	21%	29%	33%	17%
Savings/Year (in millions)	1.27	3.83	6.43	9.5
Weighted Savings/Year	0.26	1.12	2.14	1.58
Total Savings/Year	5.11			
FA WARP/Year	3.02			
Total Breakeven WARP	41.84			

Remember, it cost $1.525 million for every additional WARP in the free agent market. To find the average savings of each group, we can take the expected WARP of each group and multiply that by the cost of purchasing that WARP in the free agent market for the prospect's first six years, adjusting the FA$/WARP cost for inflation. We also know how much the prospect will cost in his first three years, and we can also find how much he will cost in his fourth-sixth years by multiplying the arbitration$/WARP by the prospect's expected WARP. We can then subtract the cost of purchasing the prospect's WARP in the free agent market by the prospect's expected cost in his first six years to determine the expected savings. Expected savings were then converted to net present value. Note that this assumes that there is steady inflation throughout baseball. This also assumes that each WARP is purchased at a fairly priced value. This also assumes that what a team purchases in WARP is what it gets. Table 3, above, shows the expected savings for a top ten hitting prospect.

Here is how to read the table. The first row shows the average WARP each subcategory produces over six years. The next row shows the chances a prospect from each subcategory is produced. The following row shows how much a team saves in millions of dollars per year if they produce a player in that subcategory. I then multiplied the savings of the subcategory by the chance of the subcategory occurring to determine a weighted savings. I summed the weighted savings to produce an average total savings per year. After that, I divided the savings by 1.69 to see how much WARP/year a team could purchase with the total savings produced. I then multiplied the savings WARP/year by the six years a team is able to control a prospect and added the average WARP of the group to come up with a total breakeven WARP. The total breakeven WARP is what a team can expect to gain in WARP from a prospect's average performance plus the additional WARP that the team could buy with the money they save from keeping the prospect. Therefore, the total breakeven WARP is what a team needs to receive in return and gain in production within six years for a trade to be beneficial, assuming that the WARP received is fairly priced. Anything above the breakeven WARP is beneficial towards the team trading the prospect while anything below the breakeven WARP is beneficial towards the team acquiring the prospect. It makes more sense to use the total breakeven WARP as the breakeven figure since prospect for prospect trades rarely happen. Tables 4, 5, and 6 show details for the other three categories.

---

[6] http://sportsline.com/mlb/salaries/avgsalaries
[7] http://mlbcontracts.blogspot.com/

Conclusion

From these tables we can see that hitting prospects have a big edge in value compared to pitching prospects. In fact, the 11-25 hitting prospects have 40% more value that the top ten pitching prospects. Top ten hitting prospects easily provide the most value of any group.

The value is high enough that it is unlikely a team could receive enough in return to trade a top hitting prospect. It also may seem that the pitching prospect breakeven figures are rather low, especially when compared to the hitting prospects. However, they do show that it is wrong to trade away a top pitching prospect for a one year or less rental, as it would be nearly impossible for one player to provide the value required in one year or less. Also, remember that these breakeven numbers do not factor in if teams are "one player away" from making the playoffs. It may be beneficial for a team to deal away a top prospect if the player it receives in return is the difference between making the playoffs and sitting at home in October. A playoff appearance can be very valuable to a team in the additional revenue it produces, especially considering that anything can happen once a team makes the playoffs. As the saying goes, flags fly forever.

It appears that teams are doing the right thing in hanging on to top hitting prospects. Trading a top hitting prospect demands a lot in return in order to ensure fair value in a trade. It also appears that teams are usually doing the right thing by not trading away top pitching prospects for a short term acquisition. There could be value to be made if a team can acquire a more certain asset it can control for over one year for a top pitching prospect, especially given the fact that even top pitching prospects are a bust over half the time. For example, if a team can acquire a player in his arbitration years, they would need less WARP in return since a player in arbitration makes less than he would on the open market.

In the end, though, it looks like teams are making the right decision when it comes to holding on to top prospects.

Victor Wang, atsbuy@yahoo.com ♦

Table 4 – Savings for Top-Ten Pitchers

	Bust	Contributor	Everyday	Star
WARP (over 6 years)	3.60	19.24	29.67	42.40
Chance of Occuring	54%	31%	11%	4%
Savings/Year (in millions)	0.61	4.11	6.45	9.29
Weighted Savings/Year	0.33	1.26	0.74	0.35
Total Savings/Year	2.69			
FA WARP/Year	1.59			
Total Breakeven WARP	22.45			

Table 5 – Savings for 11-25 Hitters

	Bust	Contributor	Everyday	Star
WARP (over 6 years)	2.25	17.54	30.35	43.40
Chance of Occuring	31%	33%	21%	14%
Savings/Year (in millions)	0.37	3.72	6.59	9.51
Weighted Savings/Year	0.12	1.22	1.41	1.36
Total Savings/Year	4.11			
FA WARP/Year	2.43			
Total Breakeven WARP	33.85			

Table 6 – Savings for 11-25 Pitchers

	Bust	Contributor	Everyday	Star
WARP (over 6 years)	2.61	18.03	30.96	44.8
Chance of Occuring	61%	24%	12%	3%
Savings/Year (in millions)	0.39	3.84	6.73	9.83
Weighted Savings/Year	0.24	0.91	0.80	0.33
Total Savings/Year	2.28			
FA WARP/Year	1.35			
Total Breakeven WARP	19.16			

Future Expectations for Overperforming Teams

Bill James

When a team over- or underperforms its Pythagorean Projection in season X, does that tell us anything about what our expectations should be for season X+1? Do teams that overperform tend to repeat their overperformance? And, if so, by how much?

Premise/Background

In late September, 2007, an issue arose on the SABR Statistical Analysis bulletin board[1] premised on the substantial overperformance of the 2007 Arizona Diamondbacks in wins vs. Pythagorean expected wins. Somebody (on or off the list, I'm not sure which) suggested that this overperformance was a fluke, and Mike Emeigh argued that the Diamondbacks overperformance might *not* be a fluke, but that "IMO, the Diamondbacks' actual performance this year is more reflective of their team quality going forward than their Pythagorean performance."

This led to an active discussion which, over the course of the next two weeks, splintered into a hundred different issues as such discussions are inclined to do. On October 13 I posted the following, as a part of one of those splinter discussions: [2]

> Well, the reason you can say that Pythagorean underperformance is a fluke is that it has a persistence of near zero. Teams that overperform in one season have basically no tendency to overperform again the next.

To which Mike Emeigh responded:

> That's true—but that's been studied to death. The question I posed is different: Will the Diamondbacks ACTUAL record NEXT year be closer to their actual record this year, or their Pythagorean record this year? IOW, are they more likely to win 90 in 2008, or 79?

To which I then responded:

> I understand that you think there's an important difference here. I don't believe there is.
>
> Take two teams, one of which scores 750 runs and allowed 800 runs and wins 76 games, as we would expect, and one of which scores 750 runs and allows 800 but wins 85 games. Numerous studies have shown that there is NO difference in their expected performance the next season ... At least this is my understanding; I haven't studied the issue myself since the 1980s, and may be behind the curve ... somebody may have found that there is some small difference in expected wins the next year ... it doesn't matter whether you are saying that they will not relapse next year because
>
> a) they will continue to outperform pythagorean expectations, or
> b) they will improve their runs scored/runs allowed ratio.
>
> because in EITHER case, you are still asserting that they have a higher expected wins next year than another team with the same runs/opposition runs ratio.   Which, as I say, you could well be right, but I'm skeptical.

I then decided that if I was going to speak publicly about this issue I should know a little more what I was talking about, so I decided to spend a few hours studying the issue.

What I am doing in this study is NOT intended to investigate Mike Emeigh's thesis that some overperforming teams are actually much better than their runs/opposition runs ratio would suggest.   To investigate that thesis would be a complicated task, and I suspect it would be impossible to refute the hypothesis given the limitations of real-world data.   I was not checking out *his* statement; rather, I was double-checking my *own* statement that "Numerous studies have shown that there is NO difference in their expected performance the next season."

---

[1] The board can be found at Yahoo! Groups, here: http://sports.groups.yahoo.com/group/StatisticalAnalysis/ It's open to all SABR members. –Ed.
[2] In the quotes that follow, typos have been corrected from the original. –Ed.

I was wondering: Is that really true, that there is no difference in their expected performance next season? Or, in saying that, was I relying on my hazy recollections of studies that I did a long time ago which might not meet the standards of modern sabermetrics anyway? Do I really *know* that there is no difference in their expected performance the next season?

And, to give away the main conclusion of the study ... I was wrong; there IS a difference.   But let me proceed in an orderly fashion.

General Method

I started with a spreadsheet containing the runs scored, runs allowed, wins, losses and games played of all teams in major league history through 2007. I will also make this spreadsheet, "Pythagoras through 2007" available to any researchers.[3]

Worksheet 1 of that spreadsheet is a simple list of the teams and data. It also includes innings pitched and "I3" (thirds of an inning), which I saved just on the off chance that it might become useful, although it never did.

I then figured the expected winning percentage for each team, using the variation of Pythagorean wins with the exponent of

$$\left(\frac{Runs + OppositionRuns}{Games} \right)^{0.286}$$

Which creates a higher exponent for teams which score and allow more runs per game, and has been shown to be more accurate that the original Pythagorean formula.

I then identified the 100 most overachieving and 100 most underachieving teams of all time in wins vs. expected wins.   Similar things were also done by other people posting to the discussion group, and their list may be as good as mine, but my list of the top overachieving teams was:

		Expected Wins	Actual Wins	Overperformance
1.	Detroit Tigers, 1905	64.87	79	+14.13
2.	Arizona Diamondbacks, 2005	65.21	77	+11.79
3.	New York Janquis, 2004	89.27	101	+11.73
4.	New York Mets, 1984	78.33	90	+11.67
5.	Kansas City A's, 1955	51.51	63	+11.49
6.	Brooklyn Dodgers, 1954	80.71	92	+11.29
7.	Cincinnati Reds, 1970	90.77	102	+11.23
8.	New York Mets, 1972	71.83	83	+11.17
9.	**Arizona Diamondbacks, 2007**	**78.90**	**90**	**+11.10**
10.	Brooklyn Dodgers, 1924	81.25	92	+10.75

The list, given in the second worksheet of the spreadsheet, ranks all teams in history 1 through 2,516, with the bottom teams being:

2512.	Pittsburgh Pirates, 1911		
2513.	Baltimore Orioles, 1967		
2514.	Pittsburgh Pirates, 1984		
2515.	Pittsburgh Pirates, 1986		
2516.	New York Mets, 1993		-14.36

My sympathies to any 96-year-old Pirate fans in the audience.

Anyway (moving on to the third spreadsheet in the worksheet) I then filled in the "next season performance" for all of these teams which had a "next season" for us to work with. (2,409 of the 2,516 teams had a next season, the exceptions being the 2007 teams, the last-season teams from the Federal League, the Players' League, the other defunct leagues, and your occasional odd pre-1903 team that went out of business.)

---

[3] At time of publication, the spreadsheet is available here: http://www.philbirnbaum.com/futureexpectations.xls . --Ed.

This enabled me to look at the average next-season performance of historically overachieving and historically underachieving teams.

The first relevant data is shown on lines 2520 and 2521 of the Worksheet 3 of this file. The "overachieving" teams in the study—100 of them—

1. Scored an average of 659 runs
2. Allowed an average of 684 runs
3. Won an average of 81.78 games
4. Lost an average of 70.35 games
5. Had an expected winning percentage of .482
6. Had an actual average winning percentage of .537 (aggregate winning percentage of .538)
7. Overachieved by an average of 8.300 games.

We have follow-up season data for 96 of these 100 teams. In the following seasons, these 100 teams again exceeded their Pythagorean expectation, but by an average of only 0.474 wins.

The "underachieving" teams in the study

1. Scored an average of 695 runs
2. Allowed an average of 683 runs
3. Won an average of 69.16 games
4. Lost an average of 84.06 games
5. Had an expected winning percentage of .511
6. Had an actual average winning percentage of .454 (aggregate wining percentage of .451)
7. Underachieved by an average of 8.68 games.

We have follow-up season data for 98 of these 100 teams. In the following seasons, these 100 teams again fell short of their Pythagorean expectation, but by an average of only 0.244 wins.

This gives us the first answer to one of the questions I wanted to study here, which was "Is the persistence of Pythagorean over achievement zero, or near zero?" It appears more likely, based on this study that it is NEAR zero, not actually zero. It appears that there is *some* persistence to the tendency to over- or underachieve vs. Pythagorean expectations, thus, that such performance is not entirely a fluke.

Targeted Method

But I'm not supposed to be reporting results yet; I'm supposed to be explaining what I did.

The real question I was trying to investigate here was "Do overperforming teams have *any* tendency to be better, in following seasons, than teams of the same runs/runs allowed ratio who are not overperformers or who are underperformers?" This was the tedious part of the study ... the rest of this stuff just took me a half-hour or something. The hard part, which consumed a few hours, was to "match" overperforming teams with very similar teams which were not overperforming.

My method was

1. To eliminate from the study any teams which had no follow-up season to be studied.
2. To take the 100 most overachieving teams of all time (remaining in the data).
3. To try to identify, for each one of those teams, a team with a nearly identical combination of games, runs scored and runs allowed, but with a very different won-lost record.

In other words, the most overachieving team of all time was the 1905 Detroit Tigers, who were outscored by 97 runs (511-608) but finished 79-74, five games over .500. They were matched with a team they actually competed with, the 1905 St. Louis Browns, who scored 508 runs (three less than the Tigers), allowed 608 runs (the same as the Tigers) but finished 54-99. (For 1905 different sources have different numbers of runs scored and allowed by these teams, although all of the data is similar to this.)

The second most-overachieving team, the 2005 Diamondbacks, scored 696 runs, allowed 856 runs, but finished 77-85. They were matched with the 2006 Tampa Bay Devil Rays, who scored 689 runs, also allowed 856, but finished 61-101.

These are near-ideal matches. Many teams do have near-ideal matches, and some teams do not. . more on that in a moment. The formula that was used to identify "matching" teams is shown in columns AL through AQ of Worksheet 4 of the file "Pythagoras through 2007". The resulting matches are shown in what would be Worksheet 5 of the file, which I have re-titled "Results".

As I said, many teams have near-ideal matches. For example, the 2005 World Champion Chicago White Sox scored 741 runs, allowed 645 runs, giving them 91.3 expected wins. They actually won 99, overachieving by 7.7 wins.

They are paired with the 1997 Los Angeles Dodgers, who

1. Played the same number of games,
2. Scored one more run (742),
3. Allowed the same number of runs (645), but
4. Finished 88-74.

But some teams do not have real good matches, and wind up being paired with teams which are similar in regard to runs scored/runs allowed per game and expected winning percentage, but pretty different in some other respect. For example, the 1894 New York Giants, who played 137 games, outscored their opponents 940-789 and finished 88-44 (+10.20 games) wind up paired with the 1930 New York Yankees, who played 154 games, outscored their opponents 1062-886 and finished 86-68 (-5.2 games). The two teams are similar in terms of runs scored per game, runs allowed per game and expected winning percentage, but the fact that the one team played 137 games including five ties and the other played 154 games makes it a less-than-ideal match. The 1936 Cardinals and the 2004 Detroit Tigers is a less than ideal match. The 1997 San Francisco Giants and the 1924 St. Louis Cardinals is a less than ideal match. The 2004 Cincinnati Reds and the 1912 Boston Braves is a less than ideal match. The 1907 Chicago Cubs and the 1972 Baltimore Orioles is probably the worst match in the study.   I'll let you study the data and reach whatever conclusion you want to about the matches.

Conclusions

In short, my statement that there is no difference in the next-season performance of overachieving and similar teams is not correct. There is a difference.   As I said, I wasn't studying the Emeigh thesis in regard to the Diamondbacks, and I can't comment on whether his hypothesis is true or not. However, based on my study, his hypothesis—which I think I can state generically as "there is a sub-set of overachieving teams which overachieve because their runs scored/runs allowed record does not reflect the true quality of the team"—appears to be plausible.

This is the specific data:

1) The 100 most overachieving teams in history (not counting those which have no next season to be studied) scored an average of 658 runs, allowed an average of 683, yet finished with an average record of 82-71 (82.02 – 70.54) consistent with previously stated data.

2) The 100 teams selected to be nearly identical to those 100 teams scored an average of 658 runs, allowed an average of 689 runs (quality leakage of six runs), but finished with an average won-lost record of 68 – 85 (68.44 – 84.72).   The overachieving teams overachieved by an average of 8.23 wins. Their "matched set" counterparts underachieved by an average of 5.09 wins, so that there was, on average, a separation of more than 13 wins between the two groups.

3) *In the following seasons, the overachieving teams outperformed their underachieving counterparts by a substantial margin.*   In the following seasons the overachieving teams scored an average of 688 runs, allowed an average of 697 runs, and finished with an aggregate total of 7,636 wins, 7,761 losses.   In the following seasons the counterpart teams scored an average of 662 runs, allowed an average of 693 runs, and finished with an aggregate total of 7,237 wins, 8,059 losses. The overperforming teams, in the following seasons, outperformed their counterparts by total of 399 wins, 298 losses—348.5 "games".

It seems to me unlikely that this is a random outcome, but I'll leave that to those of you who know those methods better than I do. I think we're in the neighborhood of four standard deviations deep.

This data and more detail about it can be found in lines 2413 and 2414 of Worksheet 4 of the spreadsheet file.

Observation

A problem with this study is that it essentially exhausts the data, making it difficult to replicate the study.   You can't study another group of 100 similarly overperforming teams, because there simply aren't any such teams.   This leaves the following options to confirm the conclusion of this study:

1) Study minor league teams (which seems to be entirely useless, because minor league teams have such high turnover that their next-season performance is not a stable or reliable measure.)

2) Look at overperforming teams in minor leagues such as the Pacific Coast League before those leagues became captive slaves of the major leagues (but it would be hard to get ENOUGH data to do that, and good luck finding the data for that.)

3) Study the same in Japanese ball (that should work if you had the data. In a few years researchers will certainly have that.)

4) Keep the 100 overperforming teams in this study, but throw out the 100 matched teams and identify 100 OTHER teams that match them fairly well (but that's would be a weak confirmation if you found the same effect or a weak rebuttal if you didn't.)

5) Replace certain elements of the study with theoretical substitutes; i.e., expected next-year performance of these teams, rather than actual next-year performance of comparable teams. (That would have been a lot easier, but I didn't do it that way because I thought it whatever I found would be less convincing.)

6) Find "best comps" for the 100 most underachieving teams, in the same way I have done for the most overachieving teams.

Bill James, biljames@aol.com ♦

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca

Does the Runs Created Formula Work for Division III Softball?

Jenn Marro and Thomas J. Pfaff

The runs-created formula for Major League Baseball accurately predicts the number of runs created by a team for MLB. Here, the authors analyze this formula for Division III softball, examine its shortcomings, and suggest a revision for the college softball game.

Introduction

The game of baseball is a strategic one where players and coaches try to outmaneuver one another, and there are many statistics that coaches use to help them with their decision making to increase their likelihood of victory. Among these statistics is the idea of Runs Created. Since creating runs ultimately makes a team successful, being able to determine which factors are more valuable in producing runs is helpful. Taking Bill James' runs-created formula and further examining it with the current statistics of Major League Baseball would be one way to determine if the factors that influence run creation have changed. However, looking at runs created in baseball and in softball shows the similarities and differences between the two games.

Softball vs. Baseball

Because the runs-created formula attempts to measure the number of runs resulting from various actions at the plate and on the base paths, we thought the formula would only need minor changes to apply to softball. To examine the concept further, we chose five Division III softball teams. The teams that were selected represent the range of abilities softball teams have, but it is important to know that they do not represent a random sample of all Division III softball teams in the country. However, we feel that the data would give accurate means for all areas of the game.

It was unclear as to which areas, if any, would have more of an influence so one-sample t-tests were used to compare the means of the five teams selected and the 2006 American League (AL) data. The AL was selected because the designated hitter rule applies in softball. Based on the numerous tests, it is clear which areas of softball were more influential and the results can be found in Table 1.

Table 1 – Comparing means of various statistics in American League and Softball

Area	AL Mean	Softball Mean	p-value
CS/AB	0.6%	1.6%	0.072
SB/AB	1.6%	4.3%	0.000
SO/AB	18.1%	18.4%	0.875
BB/AB	9.2%	10.6%	0.255
TB/AB	43.7%	37.6%	0.012
RBI/AB	13.7%	17.1%	0.045
2B/H	19.9%	13.5%	0.000
3B/H	1.8%	2.2%	0.204
HR/H	11.8%	5.5%	0.000
H/AB	27.5%	28.1%	0.669
R/AB	14.3%	27.9%	0.010
SF/AB	0.9%	0.5%	0.057
SH/AB	0.6%	2.9%	0.000
HBP/AB	1.0%	2.4%	0.047
GIDP/AB	2.4%	0.2%	0.000
Fielding %	0.984	0.885	0.000

It is also important to note here that since Major League Baseball players play 162 games as compared to softball's 40, each area was compared by percentages to take the long baseball season into account. Furthermore, the fielding percentage of softball comes from the five teams selected and not their opponents fielding percentage because the opponents of these teams overlap and would have been counted twice.

What needs to change?

Based on the results in the above table, it is clear there are differences between certain areas of softball and baseball and these should be factored into a RCF differently. Among the differences, the stolen base, sacrifice hit, fielding percentage, and caught stealing factors were considered to be the most important. Since stolen bases seemed to be a large factor in the creation of runs in softball, the stolen base version of the RCF was further examined. The Bill James stolen base version of the runs created formula is:

$$Runs = \frac{(H + BB - CS)(TB + 0.55SB)}{AB + BB}$$

Comparing this version to our five teams we have the following table:

	Predicted Runs	Actual Runs	% Difference
Team 1	253.105	325	−22.1%
Team 2	249.395	263	− 5.2%
Team 3	103.199	152	−32.1%
Team 4	178.339	228	−21.8%
Team 5	249.516	327	−23.7%
Total	1033.554	1295	−20.2%

As we can see the Bill James formula that works well for baseball is off by 20% for college softball, and always underestimates the runs created. In creating a formula for softball we recognized the need for a greater weight for stolen bases and sacrifice hits and an added component of fielding percentage. It appears that softball is very much a small ball game and these were needed to produce a better formula. Our softball formula for runs created is:

$$Runs = \frac{(H + BB + 0.1(AB - H) - CS)(TB + SB + 0.75SH)}{AB + BB + SH}$$

It is important to note here that the factor of 0.1 is used as the error rate. The error rate was taken from 1.0-FLD%, and 0.1 was used empirically as a factor since it was close to the error rate of the five teams. Hence 0.1(AB-H) was used to produce the number of times a batter got on base due to an error.

Results and Conclusions

This final formula produced relatively accurate results that can be found in the following table:

Team	Predicted Runs	Actual Runs	% Difference
Team 1	317.394	325	− 2.34%
Team 2	306.854	263	+16.67%
Team 3	134.781	152	−11.33%
Team 4	224.714	228	− 1.44%
Team 5	300.861	327	− 7.99%
Total	1284.604	1295	− 0.80%

Based on the results in the above table, a few conclusions can be drawn.

First, our formula isn't perfect. We are off by 17% for team two, which is our worst prediction and interestingly the team that was predicted best, -5%, with the Bill James version. One might conjecture that this team plays more like a baseball team than a softball team.

Second, softball is more of a small-ball game where the focus is on reaching base and advancing runners. In fact, one might suspect that the only change needed to Bill James stolen base formula would be to add the error factor of 0.1(AB-H) since there are small ball type teams in baseball and the Bill James formula works for them. But, if we do this we get a total predicted runs of 1200, which is still off by 7%. It may be that even a small-ball baseball team is still significantly different than a typical softball team. Softball steals bases at a rate almost 3 times that of MLB and has less than half the rate of home runs. So in baseball, you steal a base but are more likely to get home due to a home run, and so stealing the base didn't matter. On the other hand, in softball the stolen base is really needed to score runs. Note also that softball has 5 times the rate of sacrifice hits, which points to the high value of advancing a runner. All this leads us to believe that the 0.55 coefficient used in the baseball version and chosen empirically to fit the data for baseball isn't appropriate for softball. Lastly, we restate that this data is not a true random sample of teams, but we do feel that we obtained a good representation of different quality and styles of softball teams.

Repeating this work with a larger and truly random sample would be interesting, as well as assessing how the runs created formula works for minor league baseball and college baseball.

Jenn Marro, 235 Coddington Rd, Ithaca, NY 14850, jmarro1@gmail.com
Thomas J. Pfaff, Department of Mathematics, Ithaca College, Ithaca, NY 14850, tpfaff@ithaca.edu ◆

Working the Count
Tom Hanrahan

If a batter "works the count," forcing the starter to throw more pitches, the hitter's team could gain an advantage by forcing the pitcher out of the game sooner. Here, the author investigates just how much of a benefit that might be.

It has become fashionable in the past few years to hear players, coaches, announcers and analysts trumpet the value of "working the count". By this they mean that if the team at bat makes the current pitcher work harder by taking more pitches, prolonging many at-bats, that the pitcher may tire, and thus increase the chance of the team scoring, either off of the (less effective when tired) pitcher, or a poorer quality reliever.

What is the evidence that this postulation is true? How can we attempt to quantify the possible effect? Answering these questions is the purpose of this article.

What Might the Nature of the Advantage Be?

Possible advantages of tiring a pitcher could include

1. The pitcher may lose effectiveness past a certain point.
2. The pitcher may leave the game, and be replaced by an inferior pitcher.
3. The pitcher (in this case, a reliever) may be less able to pitch the next day.

Some have suggested that there could also be some connection between fielders becoming "bored" behind a pitcher who continues to throw pitch after pitch without retiring a batter. Studying this would require a very different approach, and so I won't address this possibility herein.

Regarding case number 1: Most managers take a starting pitcher out at a certain pitch count, either for long-term health, or because he is perceived to be ineffective in general past this point. Therefore it would be rare to so wear out a pitcher that he stays in the game while becoming much less effective than he was when he came in the game; because most managers have better options than "tired" pitchers. Additionally, we have little data to use to be able to model what this lost-effectiveness might look like. Therefore, I will discard use of #1, and instead assume a pitcher comes out (is "tired") when reaching his predetermined number of pitches. In rare cases, this could be an important issue; a championship game where the manager asks his ace to pitch when tired because no perceived option is better (Grady Little's use of Pedro Martinez in the 2003 playoffs); but this happens too infrequently to make much use of.

Case Number 2: Assume we are discussing the starting pitcher. If he throws more pitches than 'normal' to get the same number of outs, and is replaced by one of the weaker bullpen members, this could create an advantage for the duration that the weaker bullpen crew is in the game. The manager would prefer not to use this part of his bullpen; often they only come in when mopping up, or when a starter is rocked early in the game. They are typically the poorest of the team's pitchers. So, if we knew (or estimated) how much difference there is between the starting pitcher and the weak pen, and were able to determine how many extra innings they need to be used due to the offense's ability to "work the count", we could quantify this effect.

Case Number 3: The only scenario where I can see a potential significant effect here would be if a team had an ace closer in the game, who typically might not be able to pitch again tomorrow if he threw too many pitches today. In this case, if a save situation occurred the next day, the team would have to use their next-best reliever. Since the spots where this occurs would be limited to probably one inning in a small proportion of all games (save situation on back-to-back days in a series, ace closer available and much better than another reliever), I won't spend much effort on this possibility.

A variation of this could also occur in the post-season, when teams continually face each other over a week or more. Tiring out a starter in a game early in a series may make him less able to pitch on short rest later on. Again, this is a specific case that occurs infrequently, and will not be addressed here.

Measuring the Advantage of an Extra Pitch

For Case 2, the formula for the advantage of taking (or fouling off) one extra pitch can be constructed in this way.

First, define:

```
XrunsPerPitch = Extra Runs allowed per pitch taken (this is what we want to find)

QualDiff = Difference in runs per game allowed by starting pitcher versus poor bullpen pitcher
```

Then,

```
QualDiff = difference in ERA * 1.10
```

to account for unearned runs, which typically score proportionately more frequently with ERA.

```
XRunsPerPitch = extra runs per game / innings per game / pitches per batter / batters per
inning
```

In MLB 2007, the average game consisted of 39 plate appearances over 8.9 innings, which is 4.4 batters per inning. Analyst Tom Tango (a.k.a. Tangotiger) has web-published a 'pitch count estimator' which tells us that a typical at-bat lasts about 3.75 pitches. I will make more detailed use of his work later in the study.

```
XRunsPerPitch = QualDiff / 8.9 IPperG / 3.75 pitches per batter / 4.4 batters per IP

XRunsPerPitch = QualDiff / 147
```

Which is the same as

```
XrunsPerPitch = QualDiff * .0068
```

(Throughout this article, I am going to usually limit calculations to two significant figures, unless I am quoting published values. There are many assumptions that will affect the main point in much larger ways than any loss in accuracy from rounding.)

So, let's presume the existence of a starting pitcher who is much better than the weak bullpen; in this case, QualDiff might be as high as 2 runs/game. When QualDiff = 2, every extra pitch the ace starter throws adds approximately 2 * .0068 = .014 runs to the game total, merely by causing the ace's early exit and entrance of the weaker pitcher(s). A batter who finally pops out after a tough at-bat of 8 pitches has in effect bought his team .014 * 7 = a tenth of a run in terms of game strategy, over popping out on the first pitch thrown. Is this effect very large? It does not appear so, when looked at this way. Even a team of disciplined hitters who make the pitcher throw an extra 22 pitches on one day, which would send the ace to the shower about 4 outs 'early', has only increased their team's run-scoring estimate by about .31 runs; close to the value of one extra walk. If a team were to specifically pursue a strategy of taking pitches to tire out this ace pitcher, they would wind up with a net gain only if in so doing they did not in the process make any extra outs (or reduce their baserunners allowed). The advantage seems too small to risk lessening the hitters' effectiveness overall; somewhat like taking too many pitches in an attempt to ensure a runner the opportunity to steal a base, but putting the hitter in a hole. So, is there nothing to be gained? Not so fast; let's pursue a different angle.

Measuring the Advantage of Reaching Base

Instead of calculating XRunsPerPitch, let's calculate the difference in terms of reaching base versus making an out (I will call this XrunsPerOut). When a batter makes an out, the opposing team is $1/27^{th}$ closer to finishing the game. When he reaches base, the pitcher has thrown pitches to no effect in retiring the offense. If a pitcher can only throw 120 pitches in a game, every out (except for those rare 13-pitch outs!) extends how many innings he is likely to throw. Every batter who reaches base, however, adds as many pitches as his at-bat took

toward the pitcher's limit, and reduces how far the moundsman can go in the game. I will measure these differences, and compare their potential effects to the values we typically assign to batting events.

According to the aforementioned Tangotiger pitch count estimator, the average plate appearance lasts 3.75 pitches. We know in the modern game that two-thirds of all plate appearances are outs (Overall MLB on-base percentage [OBP] in 2007 was .335). Therefore, it takes an average of 3.75 divided by 2/3 = 5.6 pitches for the pitcher to achieve getting one batter out. The difference in terms of "tiring the pitcher" solely between making an out and reaching base is 5.6 pitches.

```
XRunsPerOut = XRunsPerPitch * 5.6 = (QualDiff * .0068) * 5.6
XRunsPerOut = QualDiff * .038
```

Referring back to the original example of a pitcher who was 2 runs per game better than the weak part of the bullpen (QualDiff = 2), we find that the difference between a batter making an out and reaching base would be worth 2 * .038 = .076 runs *beyond* the actual run-scoring value of the event. From this, it is easy to begin inferring that there might be some extra hidden value in reaching base.

We can use the calculations above to determine how much value there is in drawing a walk or hitting a double or striking out, in terms of tiring the pitcher and getting to a weaker reliever.

Since a batter can hit a single on the first pitch, but it takes at least 4 pitches to draw a walk, we should not be surprised some results typically take more pitches than others before they are achieved. From Tangotiger's Pitch Count Estimator, walks use an average of 5.5 pitches. Strikeouts use 4.8. All other results (hits and outs) use an average of 3.3 pitches.

Table 1 – Notional Run Value of Batting Events in Wearing out the Pitcher

Outcome	Average pitches used	Pitches subtracted by getting an out (or two!)	Total extra pitches used	Runs resulting from extra pitches used, when QualDiff = 2
Walk	5.5	---	5.5	.075
Hit (any)	3.3	---	3.3	.045
Strikeout	4.8	5.6	-0.8	-.010
Out (ball in play)	3.3	5.6	-2.3	-.031
Double Play	3.3	11.2	-7.9	-.107

Walks take 5.5 pitches, as compared to the average at-bat of 3.75; that is, an additional 1.75 pitches. But when a batter walks, and this is they key item, *he does not make an out.* The pitcher still has the same number of "outs" to get before the batter had come up; so the 5.5 pitches thrown are all "extra" pitches, beyond a typical plate appearance. The same is true for when a batter gets a hit; it "costs" the pitcher an extra 3.3 pitches on average. On the other hand, when a pitcher records an out, he has saved himself future effort, to the tune of 5.6 pitches; any out that results from fewer than 5.6 throws to the plate has saved the hurler something. Table 1 shows the value in terms of pitches for each batting outcome.

In the right-most column, I have used the previous equation XRunsPerPitch = QualDiff * .0068 to calculate the net effect of runs from the extra pitches associated with each batting outcome. We see that a walk, under the assumption made, could be worth almost .08 more runs in terms of wearing-out-the-pitcher strategy, beyond its value immediate value in creating runs; which is about one quarter of the value of a walk (.33 runs) put forward by Pete Palmer's Linear Weights in *The Hidden Game*.

So What Does It Mean?

At this point, it is tempting to add the 'hidden bonus' value of reaching base to the accepted value of each event and produce new (and improved!) values for each. The problem here is, I would be combining unlike items; "real" run-scoring values with those which are likely to produce runs in future innings. But hey, I have to come up with numerical answers somehow, don't I?

One method for quantifying this effect would be to measure this in terms of the relative worth of OBP and slugging percentage (SLG). The OBP versus SLG debate has raged for some time among sabermetricians; is OPS, the simple addition of OBP and SLG, the best way to combine them, or should OBP be weighed more heavily? Analysts have performed regressions and used other tools, and a consensus has been reached that a OBP multiplying factor close to 1.8 is an accurate estimate of correlating to runs scored (in other words, using 1.8 * OBP + SLG yields the "best" estimate of the batter's value). One unusual data point here is from Michael Lewis' book *Moneyball*; there, he wrote

that Paul DePodesta believed one extra point of OBP was worth three points of SLG. Is it possible that DePodesta was factoring in the effect of tiring the pitcher into his calculation? Well, let's see.

I will create a typical batter (call him batter A), and use Palmer's Linear Weights to assess his value. I'll then add to his OBP (more walks, fewer outs) and lower his SLG (fewer home runs, more singles), so that his Linear Weights total stays the same, and so that the total of the calculation 1.8 * OBP + SLG also stays the same (the result is batter B). Lastly, I will further adjust the batter's OBP, modifying the Linear Weights value of the batting events by the figures in Table 1 (batter C), so the modified Linear Weights is equal to batter A's.

These three batters are listed in Table 2.

[B] is equal to [A], both in terms of LW and the formula 1.8*OBP + SLG. [B]'s extra 32 pts of OBP are equally offset by his loss of 56 SLG pts.

However, when using the Mod LW values found by adding the increments in table 1, [C] is now equal to [A]. [C] compared to [A] has an advantage of 26 pts of OBP, and a disadvantage of 56 SLG

	Table 2 – Comparing batters with equal LW values										
		BB	AB	H	2B	3B	HR	LW	OBP	SLG	1.8 * OBP + SLG	Mod LW
A	50	450	115	18	3	16	1.79	.330	.416	1.010	0.33
B	70	430	110.78	10	2	10	1.79	.362	.360	1.011	2.13
C	65.6	434.4	112.40	10	2	10	0.33	.356	.360	1.001	0.33

LW (linear weights) values are walk = .33 run, single = .46, double = .80, triple = 1.02, home run = 1.40, out = -.27.

Mod LW (modified linear weights) values are those when adding in the incremental advantage of tiring the pitcher, found in table 1: walk = .405 run, single = .505, double = .845, triple = 1.065, home run = 1.445, out = -.301.

pts. Therefore, the ratio of value of OBP to SLG would have to be 56 / 26 = 2.15 to rate [A] and [C] as equals. Weighting OBP as 2.15 times as much as SLG, instead of 1.8, means an increase in the value of OBP by about 19%.

If the values in table 1 were true effects, the 'proper' ratio of valuing OBP, as it relates to SLG, would be almost 20% higher; 2.15 / 1.80 = 1.19. Not nearly enough, may I note, to fulfill Mr. DePodesta's claim of that a 3-to-1 ratio is the proper weighting.

The values in table 1 were all predicated on QualDiff = 2; the pitcher in the game was 2 runs allowed per game better than the reliever(s)-to-come-in. As this assumption changes, the change in OBP-to-SLG ratio would change, approximately proportionately.

Back to Case 3 – Tiring out the Closer

Maybe if a team can make the opponent's ace closer work hard in game 1 of a series, he will not be available in game 2. Sounds reasonable, doesn't it? Let's work through some numbers to see how large the effect might be.

Assumptions:

1. A closer who throws at least 20 pitches one day will be unavailable to pitch the next day.
2. Closers will come in for one inning (the 9th) in save situations. And in tie games when at home.
3. Based on the above, situations where closers should come in occur in 40% of all games.
4. The team's closer will "save" (or win) the game for his team 10% more often than the next best reliever. This may be a generous estimate, but we can always re-run the numbers with a lower amount.

Based on points 3 and 4, ensuring a team's closer is unavailable raises the opponent's chance of winning a game by 4% (.40 * .10) before the start of the game. This is the equivalent (assuming 10 runs typically equal one extra win, for normal situations in which non-closers pitch) of having a 0.4 run advantage. If it takes 20 pitches (5 or 6 batters) to render the closer on the shelf on day 2, then each extra pitch he throws on day 1 is worth 1/20th of the value of not having him the next day. 1/20th of .4 runs is .002 runs. So, XRunsPerPitch = .002 in the above case. This is somewhat higher than in the general case of batting against a good starter. However, it is a very limited case; only when the opponent's closer is in the game, and when you have another game schedule against the same team the next day. So, yes, there is value in

making Joe Nathan throw extra pitches. But not enough to change a manager's strategy in choosing a pinch-hitter, or enough to tell you batters to take more pitches, when the most important thing is still scoring runs *now*.

Conclusions

When facing a good starting pitcher, on a team with a weak and/or tired bullpen, there is a measurable advantage in making the pitcher throw more pitches to get the same number of outs, thus getting him out of the game earlier. This advantage can be seen in two ways:

First, for every extra pitch thrown, an extra .014 runs might score over the course of the game, in times when the starting pitcher is 2 runs per game better than the weaker part of the relief core.

Secondly, every batter reaching base (as opposed to making an out) generates a benefit of about .076 runs beyond the actual run-scoring value of the event in that inning, under the same quality assumption made above. In terms of valuing reaching base versus slugging, this means about a 20% increase assigned to the value of reaching base.

The value of a walk after a particularly long at-bat would be the combination of both of these effects.

The effect per pitch may be even higher if a team can tire out an opponent's closer, but this does not occur very often.

None of these effects are very large; not enough to justify claims that "the reason this team's lineup is so dangerous is they all really work the count"; nor enough to alter the all-time ranking of Rickey (OBP) Henderson versus Frank (SLG) Robinson, because we cannot assume the assumptions used hold over the course of a player's career.

Tom Hanrahan, Han60Man@aol.com ♦