By the Numbers

Volume 10, Number 2 The Newsletter of the SABR Statistical Analysis Committee May, 2000

Welcome

Phil Birnbaum, Editor

Welcome to the May BTN. Thanks to Rob Wood, Clifford Blau, James Fraser, and Tom Hanrahan for their time, effort, and contribution. Thanks also to the anonymous peer reviewers for your help. Hope you enjoy the issue, and please contact our contributors if you liked their work. Deadline for contributions for next issue is July 24.

A Brief Review of "Defense Independent Pitching Stats"

Clifford Blau

One of the most interesting articles I have read recently is entitled "Defense Independent Pitching Stats, " which was written by Voros McCracken. It can be found on the Internet at http://www.baseballstuff.com/fraser/articles/dips.html.

I recommend it be read by all.

To summarize this and a follow up article, Mr. McCracken states that the ratio of hits allowed by pitchers to balls in play is not a skill. Specifically, he shows that the correlation of this ratio from one year to another is so low that one cannot conclude that it is under the control of the pitcher. Initially he concluded that this hit ratio was the responsibility of the fielders, but discussion on a posting board at the same Web site led him to agree that the variation was more likely due to luck. I checked the correlation of this statistic for pitchers and hitters for the 1974 and 1975 American League, getting results similar to Mr. McCracken's. He also states that pitchers may be evaluated using an average rate of hits allowed per ball in play rather than their actual figure, as this has a higher correlation with the following season's ERA than does the current season's ERA, or component ERA (essentially ERA calculated using a runs predictor formula.)

Some comments on the article: A better test would involve a park-adjusted rate. Also, the discussion on the posting board included a consideration of career numbers. The range of this ratio using career figures in rather small, almost all pitchers being within .020 of the average. It was also determined that the career ratio is a better predictor of the following season's ratio than the current season's is. This seems to support the idea that the ratio for a season is highly dependent on luck.

Clifford Blau, 16 Lake St. #5D, White Plains, NY, 10603, brak@erols.com ♦

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
Jim Box	im.box@duke.edu	Statistics
Keith Carlson	kcarlson@stlnet.com	Economics/Econometrics/Statistics
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	HanrahanTJ@navair.navy.mil	Statistics
Keith Karcher	kckarcher@compuserve.com	General
John Matthew	john.matthew@home.com	Apostrophes
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Steve Wang	Steve.C.Wang@williams.edu	Statistics

Receive BTN by E-mail

You can help save SABR some money, and me some time, by receiving your copy of *By the Numbers* by e-mail. BTN is sent in Microsoft Word 97 format; if you don’t have Word 97, a free viewer is available at the Microsoft web site (http://support.microsoft.com/support/kb/articles/Q165/9/08.ASP).

To get on the electronic subscription list, send me (Phil Birnbaum) an e-mail at birnbaum@sympatico.ca. If you’re not sure if you can read Word 97 format, just let me know and I’ll send you this issue so you can try

If you don’t have e-mail, don’t worry–you will always be entitled to receive BTN by mail, as usual. The electronic copy is sent out two business days after the hard copy, to help ensure everyone receives it at about the same time.

E-Mailing the Editor

I have been told that my iname.com and philbirnbaum.com e-mails often don’t work. Those go through internet forwarding services before they reach me, and it seems these services are unreliable.

My “real” e-mail address – for now – is birnbaum@sympatico.ca. If that ever fails to work – who knows, I may change service providers again someday – try either phil_birnbaum@iname.com, or BTN@philbirnbaum.com.

Sorry for any inconvenience.

News on the SABR Publication Front

Neal Traven, Committee Co-Chair

One of the common threads in discussions among members of the Statistical Analysis Committee revolves around the difficulty of convincing SABR to publish statistically-oriented articles in its general-distribution journals. Though it may be somewhat self-defeating to talk about this problem in BTN, where we assuredly *do* print statistical work, a potential avenue for gaining a wider SABR audience may be at hand. In a May 17 e-mail to all research committees and regionals, SABR Publications Director Mark Alvarez floated the possibility of producing a publication composed of submissions directly from those sources. Quoting directly from Mark's email:

> The idea would be for each committee/regional to decide -- in its own way -- upon a representative article, to handle all fact-checking, photo-gathering, table-creating, etc., and to submit this finished product to me ready to roll.

I have discussed this concept with (Committee co-chair) Clem Comly and Phil Birnbaum, and we believe it to be an idea worth pursuing. But we've identified a number of issues that we'd need to deal with on the way to publication. Our thought is that the article(s) to send to Mark for this publication should be on the order of a "Best of BTN," selected from the last year of so of the newsletter. Clearly, any chosen article must be accessible to a larger audience, which might entail editing and/or rewriting the material to increase its words/technical ratio. Ideally, it would be both entertainingly informative and sabermetrically meaty. We would of course need to obtain the author's permission to republish his/her work in another publication, and most likely ask the author either to edit it or allow it to be edited.

A larger concern, however, is devising a method for selecting the article or articles to be submitted. My thought is to appoint a small panel of committee members, no more than five including Phil and either Clem or myself. This panel's charge will be to come up with a short list of worthy recent articles that can then be voted on somehow by the membership of the SAC. Any such voting would, almost of necessity, take place via email.

It's quite possible that this proposed publication won't get off the ground at all even if our research committee produces a worthy submission. It's also possible that we won't find members willing to help select articles, or won't find appropriate articles, or won't find any interest in voting on articles. But I think we should give it a shot and work toward presenting our approach to baseball research to our colleagues in the Society.

If you're interested in serving on the article-selection panel, or have a favorite article that you'd like the panel to consider, please send email to me or Phil expressing your level of interest. We expect to discuss this proposal at the SABR 30 committee meeting on June 22 – yes, an actual agenda item! – and I'd like to believe that we'll be ready to appoint the selection panel by then.

Neal Traven, 500 Market St. #11L, Portsmouth, NH, 03801; 603-430-8411; beisbol@mediaone.net ♦

Convention Meeting

The Statistical Analysis Committee Annual Meeting, at the SABR convention in West Palm Beach, will take place on

Thursday, June 22, from 3:00pm to 4:00pm.

"Baseball Prospectus" – Escaping Bill James' Shadow

James Fraser

The "Baseball Prospectus" annual is the latest evolution of the Bill James Baseball Abstract. While not nearly as absorbing as James' work, writes the author, it does cover baseball with depth, intelligence, and solid sabermetric methods.

Baseball Prospectus 2000 begins with a forward from current ESPN.com columnist, and former Bill James assistant, Rob Neyer. Neyer's introduction sets the tone for the entire book: *BP* is the product of people who were greatly influenced by the father of sabermetrics, a group that is growing in size and influence. *Baseball Prospectus* is the next generation of *Bill James' Baseball Abstract*s, and will play a pivotal role in increasing the awareness of sabermetrics. While the book owes much to James, it does manage to craft out a unique identity. *Baseball Prospectus* is structured differently, there are a greater variety of voices, and although the essays are not as absorbing as James', *BP* is definitely worth a read.

Baseball Prospectus 2000

By Chris Kahrl, Clay Davenport, Joseph S. Sheenan, Rany Jazayerli and a Team of Experts on Baseball Talent

Brassey's Sports, 536 pages (paperback), $21.95

ISBN 1-57488-214-7

The skeleton of *BP* is Clay Davenport's complex, yet excellent statistical analysis. His Davenport Translations (like James' Major League Equivalencies), and Equivalent Average system (like James' Runs Created per 27 outs, but made to fit the batting average scale) are explained in a friendly question and answer format. The Q and A allows *BP*'s extremely statistical methods to be explained without intimidating the large portion of their audience that isn't interested in the science behind sabermetrics, but is interested only the results of the analysis. For those interested more in the mechanism behind the book's system, your best bet is its excellent web site (www.baseballprospectus.com).

The greatest strength of *BP* lies in the ability of the Davenport Translations to compare every single player in organized. Using a series of "difficulty factors," which are regrettably not published anywhere in the book, the Translations can be used to directly compare players in A-ball (Rafael Furcal or any other hot prospect) to players in the majors to players in Japan (*BP* stubbornly includes Roberto Petagine). Pitchers also have their statistics translated, with a focus on peripheral statistics. The almost 2000 players covered in the book are presented with these Davenport Translations and brief, information-packed comments. The table below shows some selected translated EqA's; remember, these ratings take into account league and park.

Each player capsule also includes a projection line, created by Propsectus's "Wilton" projection tool. Exactly what Wilton is or how it works is not answered in the book, but to the best of my knowledge it uses historical comparisons and a "neural net" system to derive projections. *BP* hasn't tried to project pitchers, for good reason, especially given the number of minor league pitchers covered. Wilton is an interesting tool, but far from revolutionary. I'm sure that some people really value the projections, but for me they're only a sideshow.

The best, and usually longest, discussions are reserved for the most interesting players, but all players receive sharp, often sardonic, always amusing comments. The breadth of their comments and translations -- they truly live up to their claim of covering A-ball to the majors -- is a valuable asset for learning about future stars and gaining new insights into veterans.

Selected EqA's

Name	Team	League	EqA
Mark McGwire	STL	NL MLB	0.344
Rey Ordonez	NYM	NL MLB	0.220
Nick Johnson	NOR	EAS AA	0.340
Rafael Furcal	MAC	SAL A	0.262
Roberto Petagine	YAK	JAP	0.325

The player comments are organized by team and preceded by a team essay. The team essays are rather uneven. Some simply recap the 1999 season and rant about the stupidity of the local general manager, while others raise interesting points of discussion. Although all are written in good humour, I prefer those that break the mold of: This is where the team was, this is where they are and this is where they're going. My favourite was the analysis of the Cincinnati Reds, including a fascinating discussion of their unconventional bullpen usage. Contrasting the relief usage patterns of the Reds with the Indians makes a compelling argument for using your best reliever when the game situation requires it, not when Jerome Holtzman's save rule says the manager should reach for the phone.

Three other systems independent of the Davenport Translations are used to evaluate pitchers: Pitcher Abuse Points, Support-Neutral Records and a new Reliever Evaluation Tool. Rany Jazayerli and Keith Woolner revisit last year's biggest hit: Pitcher Abuse Points (PAP). The system, far from perfect, provides insight into pitcher workload and especially the danger of overworking young pitchers. PAP needs improvement; right now it's nothing more than a surprisingly revealing system that is not based in empirical evidence. The authors promise to improve the system using historical data, but I'm not entirely sure whether PAP will ever be anything more than a system of educated guesses.

Mike Wolverton takes advantage of more detailed game data to analyze both starting and relief pitching. His Support-Neutral Records system tells us what a pitcher's won/loss record would have been given average run support. The pitcher is assigned a certain number of wins and losses for each game he appears in, based on the base/out situation when he left the game. I've never understood the strange fascination with giving a pitcher a win or a loss, but I suppose that Wolverton's is an improvement over figuring after the fact support-neutral records, and a massive improvement over traditional won/loss records. Wolverton also contributes a new reliever evaluation tool; it is an update of an old *STATS Scoreboard* standby called "runs prevented", with the addition of park factors and some other slight technical tweaks. The reliever evaluation tool is information worth having, but doesn't really deserve the almost 20 pages it takes up.

There are some other nice features that I really enjoyed. Rany Jazayerli chimes in with a top 40 prospects list, and Keith Woolner presents a list of great questions for sabermetrics to answer in the next millennium. Both features are enjoyable, and I hope that *BP* makes an attempt to answer some of Woolner's questions in the future.

This is my second *Baseball Prospectus*, and it is a book I will continue to buy in the future. The strength of the book is the Davenport Translation system, which allows for comparison across levels. The writing will probably never be as good as Bill James', but I think that *Baseball Prospectus* has its own identity and need not remain in James' shadow. They cover a far broader range of players, have more involved statistical analysis and utilize a wide variety of voices. The format of *BP* will be similar next year, they'll introduce new tools, improve some old ones. *Baseball Prospectus* is a worthy buy both for the amount of players it covers and the sabermetric sophistication with which it covers them.

James Fraser, 297 Old Orchard Gr., Toronto, Ontario, Canada, M5M 2E6, jfraser99@hotmail.com. ♦

Book Reviews Wanted

Every year, a number of books and magazines are published with a Sabermetric slant. Many of our members have never heard of them. Our committee members would like very much to hear when this kind of stuff comes out.

If you own a copy of any baseball book of interest, we'd welcome a summary or a full-length review. The only restriction, please: the book should have, or claim to have, some Sabermetric content.

The James Fraser review you've just finished reading on this page is exactly the kind of thing we're looking for.

Send reviews to the usual place (see "Submissions" elsewhere in this issue). Drop me a line if you want to make sure no other member is reviewing the same publication, although multiple reviews of the same book are welcome, particularly for major works. Let me know which book you're doing, so I don't assign the same book twice.

And if you're an author, and you'd like to offer a review copy, let me know – I'll find you a willing reviewer.

"Best Teams" Logic Flawed

Clifford Blau

Is it harder to compile an outstanding record against teams of equal mediocrity, or against a combination of extremely bad and extremely good teams? An internet-published article says the former, but, here, a study by the author suggests the latter.

During the 1998 season, James Kushner published an article entitled, "The Best Teams in Baseball History." (It is available at the Baseball Prospectus Web site, http://www.baseballprospectus.com/news/19980728kushner.html.) In it, he attempted to determine which team was best essentially by comparing the records of all first place teams to the standard deviation of the records of the other teams in their leagues. As he pointed out in the article, competitive balance has tended to increase throughout major league history, and a .670 record was as hard to achieve in the 1990's as a .800 record was in the 1880's. By using the standard deviation of the opponents' records, he believed that he was accounting for the varying strength of the competition. His premise was that it is harder to achieve an outstanding record in a league with a lot of mediocre teams than it is in a league with a couple of good teams and a bunch of patsies.

This premise has bothered me ever since I read the article. While the quality of the opposition clearly affects the difficulty of achieving an outstanding record, can you measure that quality simply using standard deviations? The 1906 Cubs were not one of the best teams ever, according to Mr. Kushner, because they "had two pathetic teams to beat up on (the Cardinals and Braves were both under .400), and two other teams did break .600, showing that winning wasn't difficult (or, at any rate, unique)." On the other hand, he rated the 1941 Yankees as the best ever, in part, because they played in a league with 7 teams whose records weren't far from .500. I believe that the final paragraph of Mr. Kushner's article reveals the weakness of his method. In it, he states that the Yankees team of 1998 to that point ranked fourth best of all time. Furthermore, if the last place Devil Rays improved their record over the remainder of the season while the second place Red Sox slipped, the Yankees had the opportunity to be the best team ever. Reading that, I asked myself, how would Boston losing a series of games to Tampa make the Yankees a better team? Obviously, it wouldn't.

In 1906, the Cubs had a winning average of .763 and the other NL clubs had a collective record of .462. However arranged, those clubs would have to have had a collective record of .462. If the Giants and Pirates had been weaker, and the Braves and Cardinals stronger, would it have been harder for the Cubs to win 116 games? To answer this question, I ran a simulation of the 1906 season using a computer game (Baseball for Windows 1.0). First I played twenty seasons with the original teams. The Cubs won between 94 and 118 games each season, with an average record of .711. The standard deviation of the other seven teams' record was .121. I then evened out the competition by switching the Giants' pitching staff with the Cardinals', and the Pirates' nonpitchers with those of the Braves. With these new teams, I played another twenty seasons. This time, the Cubs' record averaged .723, with their wins ranging from 103 to 121. The other seven teams had a standard deviation of .031. While this is far from conclusive, it certainly doesn't support the idea that a balanced league makes it tougher to win a lot of games.

This simulation points out another weakness of ranking teams solely on wins. Luck plays a big role in the number of games a team wins in a single season. Most of the teams with the best records in history were probably lucky as well as good. It is probably better to use other factors, such as runs scored and allowed, in rating teams, as well as to examine the teams for a period of at least three seasons.

A counterargument might be that a less drastic reordering of talent between the teams might lead to an improvement in the strength of the teams. This could occur if the good teams had infrequently used players who were superior to regulars on the poor teams. For example, if Art Hofman had been on the Cardinals instead of warming the Cubs' bench, the overall level of talent in the league would have been increased. Also, a great team in a balanced league may not achieve its best possible record without any strong competitors to push it.

Clifford Blau, 16 Lake St. #5D, White Plains, NY, 10603, brak@erols.com ♦

Probability of Performance – A Comment

Rob Wood

In a previous BTN, Mike Sluss introduced "Probability of Performance," a statistic that represents the chance of a league-average player achieving an outstanding result. Here, the author suggests that changing the baseline to a random player, instead of an average player, gives a more realistic and meaningful result.

Introduction

Mike Sluss (BTN November 1999, BRJ 1999) recently introduced a relatively new method to baseball analysis. "Probability of Performance" (POP) is designed to help us determine which of two achievements is more outstanding. For example which is the more outstanding batting average: Carl Yastrzemski hitting .301 in 1968 when the AL hit .230, or Bill Terry hitting .401 in 1930 when the NL hit .303?

Using standard binomial formulas, Sluss calculates the probability that a batter hitting at the league average pace would have a batting average at least as high as the league leader's batting average. For ease in presentation, Mike converts the probability to an inverse logarithmic scale to arrive at POP. According to Mike's method, all that is required is the league leader's number of hits and at bats and the league-average batting average (excluding the league leader).

Following the custom of removing pitchers' at bats, I calculate that Terry's POP of 6.13 exceeds Yaz's POP of 3.36, so Terry's achievement is deemed more outstanding. How do we interpret these numbers? By undoing the logarithmic transformation, Terry's 6.13 POP means that the probability of his achievement is roughly 1 in 1.3 million, since the probability of a 1930 NL league average batter hitting at least .401 is roughly 1 in 1.3 million, using the standard binomial formulas. Yaz's POP of 3.36, for comparison, means that the probability of an AL 1968 league average batter hitting at least .301 is only 1 in 2,300.

Well, these probabilities are much smaller than I would have guessed. After thinking about it for a while[1], I realized that these numbers were correctly calculated from Mike's formulas. The problem is that I (mistakenly) thought POP was designed to answer the following type of question. What is the likelihood that a 1968 AL batter would hit at least .301? What is the likelihood that a 1930 NL batter would hit at least .401?

If this were the "event" for which the probability was to be estimated, I guessed that we would get much larger probabilities. I asked myself whether it really is a 1 in a million chance that there would be a .400 hitter in a league that hits over .300? After all, in the NL 1930 Babe Herman hit .393, Chuck Klein hit .386, and Lefty O'Doul hit .383. And Lajoie, Hornsby, Sisler, Cobb, Jackson, Williams, and Heilmann also hit over .400 in other years. Terry's .401 average is not so spectacular compared to these other high averages.

In this article, I want to follow through on my initial concerns about POP. I will attempt to discover why POP may not be the best or only measure of the degree of outstanding achievements, at least not in the way that I think of "outstanding achievements." Following the investigation, I will introduce a revised POP formula that provides more reasonable estimates of the likelihood of outstanding achievements (in my view).

Stylized Example

Before I get back to the real-world of baseball and its statistics, consider the following stylized example of two different leagues. The following two tables will present all the relevant data along with Mike Sluss's POP results (labeled MS-POP). In addition, I will present the results of my own revised POP formula (labeled RW-POP) for comparison.

[1] I also benefited from a lengthy and provocative email dialogue with editor Phil Birnbaum; in addition I want to thank peer reviewers who provided comments on a draft of this article.

Player A, the leader in League 1, hit .310 in a league that hit .254 (removing his own at bats). This leads to an MS-POP of 2.56, indicating a 1 in 360 chance. My own formula, described below, RW-POP is 2.54 in this league, or a 1 in 350 chance. Let's consider MS-POP and RW-POP to be equivalent in this league.

League 1

Player	H	AB	Avg
A	155	500	.310
B	128	500	.256
C	127	500	.254
D	126	500	.252

Player F, the leader in League 2, hit .308 in a league that hit .240 (removing his own at bats). This leads to an MS-POP of 3.50, indicating a 1 in 3,200 chance. Do we really believe that player F's achievement is that rare? (After all, player G did hit .300.) By comparison, RW-POP is 0.92 in this league, indicating a 1 in 8 chance.

League 2

Player	H	AB	Avg
F	154	500	.308
G	150	500	.300
H	120	500	.240
J	90	500	.180

To recapitulate, MS-POP is significantly higher in league 2 than in league 1 and therefore deems player F to be more outstanding than player A. On the other hand, RW-POP is significantly higher in league 1 than in league 2, and therefore deems player A to be more outstanding than player F. Which is "right"?

Please take a look again at the two leagues. I would guess that 99 out of 100 people would say that player A's achievement is more outstanding than player F's. In league 1, player A is head and shoulders above everyone else in the league, whereas in league 2, there is a much greater diversity of performance and indeed player G does just about as well as player F.

Another way to say the same thing is to ask which method gives more reasonable answers. Looking at the data for players G, H, and J in league 2, what would you think is the likelihood of player F hitting .308? Do you think it around 1 in 3,200 or around 1 in 8? I think most people would say around 1 in 8. In any event, I am sure nearly everyone would say that player A had the more outstanding achievement compared to player F, and a method that claims otherwise needs to be looked into.

So what is going on? MS-POP only looks at the league average and ignores the spread and all the rest of the players. Thus, MS-POP is more impressed by player F in league 2 (the low average, high standard deviation league) than player A in league 1 (the high average, low standard deviation league). In my opinion, this can lead to results that are at odds with what we commonly mean by an "outstanding achievement."

Analysts have wrestled with these issues for years. David Shoebotham, Merritt Clifton, and Ward Larkin, separately in the 1970's and 1980's, developed "relative" baseball methods that we still use today. They considered several different methods. One is to compare the league leader to the league average (essentially Mike Sluss's method); another is to compare the league leader to the runners-up; and others attempt to take into account the spread (standard deviation) among all players. Following in this tradition, I will introduce a version of POP that takes into account the entire distribution over all players.

Revised POP (RW-POP)

By reconsidering the original intent of POP, I will revise the MS-POP formula. We will see that the new formula (RW-POP) does not suffer from any of the concerns mentioned above. There is a price to pay, however. MS-POP is fairly easy to calculate. All that is needed is a league average and the league leader's performance. RW-POP, on the other hand, will require the performances of each player in the league; today, though, this data is widely available in electronic form.

Let me characterize the original intent of POP as a desire to answer the following question. What is the likelihood that a player will achieve the league leader's batting average? Mike Sluss operationalized this question by asking: "what is the likelihood that a player hitting *at the league average pace* would achieve the league leader's batting average?"

I wish to argue that this operationalization is where POP is open to question. By considering the hypothetical league average batter, it is easy to see that POP's probabilities will be very small. It's like asking: "what is the likelihood of the league average batter leading the league?" Very small indeed.

I will operationalize my characterization of the original intent of POP by asking a different question. What is the likelihood that a player *chosen at random* would achieve the league leader's batting average? This question seems to me to be more directly related to the original intent of POP.[2]

For comparison to what will follow, I present Mike Sluss's original formulas:

$$\text{MS-POP}_{prob} = \sum_{i=H_M}^{AB_M} \frac{AB_M!}{(i!)(AB_M-i)!}(Avg_{L'})^i(1-Avg_{L'})^{AB_M-i}$$

$$\text{MS-POP} = -\log_{10}(\text{MS-POP}_{prob})$$

where H is hits, AB is at bats, the subscript L′ refers to the entire league (excluding the league leader), M refers to the league leader, and "Avg" refers to batting average. Of course, POP can be calculated for any "rate" variable in baseball such as batting average, on base average, walk rate, home run rate, etc.; I suppose POP could also be applied to slugging percentage, though SLG is not truly a rate.

The formulas for RW-POP, given below, are straightforward extensions of the MS-POP formulas. In words, I calculate the likelihood of each player in the league achieving the league leader's batting average, and then weight these likelihoods by the player's at bats. I then convert over to the inverse log 10 scale for ease in presentation, though this is less necessary now since the probabilities are more reasonable. (If anyone is interested, I would be happy to pass on my APL programs.)

$$\text{RW-POP}_{prob} = \sum_{j \neq M} \frac{AB_j}{AB_{L'}} \sum_{i=H_M}^{AB_M} \frac{AB_M!}{(i!)(AB_M-i)!}(Avg_j)^i(1-Avg_j)^{AB_M-i}$$

$$\text{RW-POP} = -\log_{10}(\text{RW-POP}_{prob})$$

To get a sense of how this works, let's walk through how MS-POP and RW-POP are calculated for stylized league 2 above. The likelihood that player G (the .300 hitter) would hit at least .308 in 500 at bats is .3642; the likelihood that player H (the .240 hitter) would hit at least .308 is .0003; and the likelihood that player J (the .180 hitter) would hit at least .308 is virtually nil. Taking the weighted average of these probabilities, weighting by the player's at bats, yields the RW-POP prob of .1215 (about 1 in 8).

On the other hand, MS-POP calculates that the likelihood that a batter hitting at the league average pace (.240) would hit at least .308 in 500 at bats is .0003, or about 1 in 3,200.

MS-POP vs. RW-POP

The reason why MS-POP and RW-POP give different answers is due to inherent non-linearities. As seen above, the binomial likelihood of achieving a specific batting average is a highly non-linear function of the assumed fixed pace. The likelihoods to hit .308 presented above were seen to be .3642, .0003, and 0, for the hitters who had .300, .240, and .180 averages, respectively.

Since MS-POP calculates the likelihood using the average average, MS-POP's probabilities will always by much less than the average probability considering each player's likelihood (as done by RW-POP). And when converted to the inverse log 10 scale, MS-POP will always be much greater than RW-POP. Colloquially, MS-POP considers the *average* player and calculates the likelihood of the average, whereas RW-POP considers the *random* player and calculates the average likelihood.

[2] The crux of the matter is the difference in our definitions of an "outstanding achievement". Unfortunately, there does not appear to be any objective way to determine the proper definition of such a slippery concept.

As stated above, I believe that at their heart MS-POP and RW-POP are designed to answer the same question. However, they go about it somewhat differently since they make different (implicit) assumptions. An unavoidable task when we ask ourselves how rare an event is is to determine the "alternative" outcomes. In our case, when we are interested in estimating how rare is the achievement of the league leading batter, we must ask ourselves what we "expected" the league leader to have hit, before we knew he actually led the league.[3]

My view is that MS-POP implicitly assumes that the "prior" probability distribution for the league leader's performance is given simply by the league average. This is tantamount to assuming that all players in the league have equal abilities. To be clear, I am not saying that Mike Sluss actually believes this; however, his formula would provide the "correct" answer if players were all of equal ability.

On the other hand, RW-POP implicitly assumes that the "prior" probability distribution for the league leader's performances is given by the other players' performances in the league. This is tantamount to assuming that batting averages reflect true abilities in the league's players.[4]

While it can be argued that both assumptions are rather extreme, I think almost everyone would agree that the assumption underlying RW-POP is much more reasonable. After all, we do not really believe that Tony Gwynn and Mario Mendoza are equally good hitters, do we? It is much more reasonable to assume that there is a distribution of hitters' batting averages throughout the league, and this distribution (not just the average) provides the context in which to properly measure achievement.[5]

The stylized example given above showed that this distribution can affect how "outstanding" we deem an achievement. Accordingly, a formula such as RW-POP that takes into account this distribution may often produce a more reasonable result than a formula such as MS-POP that does not.

Terry vs. Yaz

Now that we have a new toy, let's try it out by returning to Bill Terry and Carl Yastrzemski. This duo exemplifies the need to consider performance in the proper context. Recall that we would like to compare Terry's .401 average in NL 1930 to Yaz's .301 average in AL 1968. Given that non-pitchers hit .312 in NL 1930 as compared to .238 in AL 1968, which was the more outstanding achievement?

According to MS-POP, Terry's achievement (MS-POP of 6.13) was a 1 in 1.3 million chance. Yaz's achievement (MS-POP of 3.36) was a 1 in 2,300 chance. By this measure Terry's achievement was far more outstanding, approximately 600 times more unlikely, than Yaz's. I began the article by questioning these results, since they are at such odds with my intuition. Recall that Shoebotham recommended the ratio of the league leader's average to the league average. Doing that here shows Terry to be 28.5% better than league average, and Yaz 26.5% better. This result matches my intuition much better than the results of MS-POP given above.[6]

Let's see how RW-POP sees Terry vs. Yaz. According to RW-POP, Terry's achievement (RW-POP of 1.72) was a 1 in 52 chance. Yaz's achievement (RW-POP of 1.47) was a 1 in 30 chance. As I said above, don't these likelihoods seem more "reasonable" than a 1 in a million probability?[7] Even by RW-POP, we still see that Terry's achievement is deemed to have been more outstanding than Yaz's, but is now considered only about 2 times more unlikely.

The large differences between MS-POP and RW-POP make me guess that the all-time seasonal and career rankings by the two measures are undoubtedly different. It would be interesting to see how the two methods compare and contrast throughout baseball history.

---

[3] In Bayesian statistics terminology, this is the "prior" probability distribution.

[4] This assumption introduces a bias into RW-POP since players who have had good years are likely to have been "lucky" and performed above their own "prior" expected levels. Per Phil Birnbaum's suggestion, I ran a simulation of 100 leagues each having 100 players who each get 500 at bats using a reasonable distribution for "true" batting averages. The average RW-POP using true batting averages is 1.44, compared to the average RW-POP using the observed batting averages of 1.86. (Using the same data, the average true MS-POP was 4.19 and the average MS-POP using observed batting averages is 4.24.) This bias in the RW-POP calculation is not likely to vary much over time, and therefore is unlikely to affect cross-era comparisons using RW-POP, and in any event would not be easy to eradicate. In comparison to the difference between MS-POP and RW-POP, the bias in RW-POP is small.

[5] Stated another way, Mike Sluss's version of POP implicitly attributes *all* the difference between the league average and the league leader's average to luck; whereas my method implicitly assumes that *none* of this difference is due to luck.

[6] In addition, according to my back-of-the-envelope estimate, Terry and Yaz each were roughly 3 standard deviations above their league average.

[7] To be clear, MS-POP is a valid probability and I am not questioning its underlying math. I am questioning whether the event for which it is calculated, the league average hitter leading the league, is a particularly meaningful event in baseball terms.

Concluding Remarks

In this article, I have raised concerns about Mike Sluss's Probability of Performance (MS-POP) measure. Mike has made a significant contribution to baseball analysis. His concept is great. I have argued, however, that his operationalization of the concept may be open to debate and alternative renderings. MS-POP considers only the league average batting average, and thereby narrowly defines an "outstanding achievement."

By considering all players in the league, not just the league average player, I have developed a revised version called RW-POP. It requires data on each player in the league; however, this data is readily available (e.g., via Sean Lahman's website). RW-POP asks what is the likelihood that each player in the league would achieve the league leading performance, and then weights these likelihoods by the player's at bats. In this way, the formula can take into account the diversity of the performances, including the outstanding performances of runners-up. Accordingly, RW-POP has a slightly broader definition of an "outstanding achievement."

I have shown that the probabilities calculated via RW-POP are more "reasonable" than those calculated via MS-POP. In addition, a stylized example has shown that RW-POP can be more consistent with our common sense definition of an "outstanding achievement" than MS-POP.

POP is a great idea and I hope it takes hold in baseball research. As such, I think it merits serious contemplation and investigation. I hope I have begun that process in this article.

Rob Wood is a management consultant in California. Rob Wood, 2101 California St. #224, Mountain View, CA, 94040-1686, rob.wood@us.pwcglobal.com. ♦

Submissions

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 (but no death threats, please) are all welcome.

Articles should be submitted in electronic form, either by e-mail or on PC-readable floppy disk. 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).

I usually edit for spelling and grammar. (But if you want to make my life a bit easier: please, use two spaces after the period in a sentence. Everything else is pretty easy to fix.)

Deadlines: January 24, April 24, July 24, and October 24, for issues of February, May, August, and November, respectively.

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

Clutch Teams in 1999

Tom Hanrahan

Long-standing sabermetric results have shown that (a) teams score predictable numbers of runs given their offensive batting line, and (b) teams win a predictable number of games given their runs scored and allowed. Teams that overshoot or undershoot their runs or wins estimates can be said to have been good or poor in the clutch. Here, the author examines which teams those were in 1999.

Which teams came through in the clutch last season, exceeding the expectation of victories they might have had based on their statistics? Which teams couldn't buy a hit at the right time, and consequently had a more disappointing year than they might have had? This article attempts to answer these questions.

There are two basic premises on which this analysis is based:

1. The number of games a team will win can be predicted accurately by the number of runs they score and allow.
2. The number of runs a team will score can be predicted accurately when the team statistics of hits, walks, total bases, and steals/caught stealing are known.

Both of these have been shown to be undeniably true by many students of the game.

Predicting Wins

A well-known formula to predict expected wins, sometimes called the Pythagorean formula, is

$$\text{Expected Wins} = \text{Games Played} * \text{runs scored}^2 / (\text{runs scored}^2 + \text{runs allowed}^2)$$

So, if a team wins more games than they "should" have by this formula, we can say they did well in what I will call the "game clutch" category. They likely got many timely hits in close games, and maybe their bullpen was effective in "save" situations.

Predicting Runs

There are also common equations used to predict runs scored, many of these developed and/or popularized by Bill James. A version of the "runs created" formula is

$$\text{Expected Runs} = (\text{Hits} + \text{Walks} - \text{Caught Stealing}) * (\text{Total Bases} + .52 * \text{Stolen Bases}) / (\text{At Bats} + \text{Walks})$$

There are more complicated (and slightly more accurate) versions of this formula, but I chose to use this one because of the ease of finding and manipulating data for 1999.

Testing

If a team scores more runs than the formula predicts, they did well in the "run clutch" category. They likely batted well with runners in scoring position, and cranked their home runs with men on instead of hitting solo shots.

We can measure how many extra games each team won (or lost) by being game clutch directly from the formula, Actual Wins minus Expected Wins. We can measure extra games won by being run clutch by finding the difference in expected wins when using the Expected Runs instead of actual runs scored in the Expected Wins formula.

The chart on the next page gives each team's offensive statistics, and their extra wins from run clutch, game clutch, and both clutch measures combined.

Observations

There is more variation (as measured by the standard deviation) in the game clutch than the run clutch measurement. In other words, last year's results were altered more by hitting well when the game is on the line than by hitting well with runners on base.

By far the worst game clutch team was the Royals. They scored many more runs than the Cubs and Devil Rays, and allowed about the same amount, and finished behind them both. The Royals' deficiencies in their bullpen last year (more Blown Saves than Saves!) have been loudly trumpeted by many as the cause of their long year.

The Orioles were the worst American League team in run clutch performance, scoring 32 less runs than they "should" have. Additionally, they were the 2nd worst team in the majors to the Royals in game clutch, winning only 78 games in spite of easily outscoring the opposition. The table shows the Orioles could have been expected to win 9 more games last season than they actually did win. Those 9 games would have made them the 5th best team in the league, with a record of 87-75. Their team batting average, on base average, and slugging percentage were all better than the Oakland A's', they had more stolen bases, and their pitching and defense allowed fewer runs, yet somehow they managed to win 5 fewer games. The A's run at the West title was truly a remarkable one.

There were no significant winners or losers among National League clubs, but the Astros' "extra" game they won did allow them to steal the division title from the Reds, whose "extra" loss cost them a playoff appearance.

American League

	AB	H	TB	W	SB	CS	RC	RUNS	RA	WINS	extra wins, run clutch	extra wins, game clutch	extra wins, total clutch
NY	5568	1568	2521	718	104	57	913	900	731	98	-0.7	0.4	-0.3
Bos	5579	1551	2497	597	67	39	865	836	718	94	-2.2	0.8	-1.4
Clev	5634	1629	2629	743	147	50	985	1009	860	97	2.3	3.2	5.5
Tex	5651	1653	2705	611	111	53	975	945	859	95	-2.1	6.3	4.2
Bal	5633	1572	2522	615	107	46	883	851	815	78	-2.6	-6.5	-9.1
Tor	5642	1581	2582	578	119	48	897	883	862	84	-0.9	1.1	0.2
Oak	5519	1430	2462	770	70	37	859	893	846	87	3.6	1.6	5.2
Sea	5572	1499	2536	610	130	45	869	859	905	79	-0.5	2.2	1.7
Chi	5644	1563	2421	499	110	50	812	777	870	75.5	-3.1	3.6	0.6
Det	5481	1433	2426	458	108	70	761	747	882	69.5	-1.0	1.8	0.8
KC	5624	1584	2435	535	127	39	845	856	921	64.5	1.5	-10.6	-9.1
TB	5586	1531	2296	544	73	49	771	772	913	69	0.5	1.5	2.0
Ana	5494	1404	2170	511	71	45	687	711	826	70	3.1	1.1	4.2
Min	5495	1450	2110	500	118	60	685	686	845	64	0.6	-0.4	0.3
						AVGS	843	838	847	80.3	-0.1	0.4	0.3
						S DEV	92	90	60	12.2	2.1	4.2	4.5

National League

	AB	H	TB	W	SB	CS	RC	RUNS	RA	WINS	extra wins, run clutch	extra wins, game clutch	extra wins, total clutch
Ari	5658	1566	2595	588	137	39	903	908	676	100	0.9	-4.2	-3.4
Atl	5569	1481	2427	608	148	66	820	840	661	103	2.3	3.0	5.3
Cin	5621	1534	2546	565	164	54	870	865	711	95.5	0.0	-1.2	-1.2
NY	5539	1544	2404	709	150	61	871	853	711	96.5	-1.2	0.9	-0.2
Hou	5485	1463	2306	728	166	75	815	823	675	97	1.2	0.1	1.4
SF	5564	1507	2413	696	109	56	847	872	831	86	2.8	1.1	3.9
LA	5567	1480	2340	595	167	68	790	793	787	77	0.7	-4.6	-3.9
Phi	5598	1539	2412	631	125	35	849	841	846	77	-0.3	-3.5	-3.8
Pit	5468	1417	2292	573	112	44	757	775	782	78.5	2.3	-1.8	0.6
StL	5570	1461	2371	613	134	49	799	809	838	75.5	1.4	-2.6	-1.2
SD	5394	1360	2218	631	174	67	737	710	781	74	-2.6	0.7	-1.9
Mil	5582	1524	2378	658	81	33	833	815	886	74.5	-1.4	0.3	-1.1
Col	5717	1644	2696	508	70	43	926	906	1028	72	-1.3	1.2	-0.1
Mon	5559	1473	2376	438	71	51	748	718	853	68	-2.8	0.8	-2.0
Flo	5578	1465	2203	479	92	46	705	691	852	64	-1.1	-0.3	-1.4
Chi	5482	1411	2303	571	60	44	747	747	920	67	0.4	2.6	3.0
						AVGS	814	810	802	81.6	-0.1	-0.4	-0.6
						S DEV	63	67	100	12.8	1.7	2.0	2.3
	BOTH LEAGUES					AVGS	828	823	823	81.0	0.0	0.0	0.0
	COMBINED					S DEV	78	79	86	12.4	1.9	3.3	3.6

Notes:

1. Abbreviations: TB is total bases, RC is Runs Created (= Expected Runs), RA is runs allowed
1. The teams were put in order of best to worst ratio of runs scored to runs allowed
1. The average team scored 5 less runs in 1999 than predicted by the formula, and so 5 runs were added to the run clutch equation to force the average to zero.

Tom Hanrahan, 21700 Galatea St., Lexington Park, MD, 20653, HanrahanTJ@navair.navy.mil. ♦

How Often Does the Best Team Win the Pennant?

Rob Wood

The best team in the league by talent may not be the team that goes on to the World Series. Through its own bad luck, or another team's good luck, the most talented team may find itself knocked out of the pennant race. How often does this happen? Here, the author designs a simulation to illuminate the issue.

Introduction

I have frequently wondered how good a job baseball does in identifying the best team in the league. After all, there is a significant amount of "luck" inherent in baseball. A line drive can be just foul by inches, a ground ball can be just barely out of the reach of a fielder, many pitches are borderline balls/strikes, etc. Countless games can go either way during a season. 154 or 162 game seasons may not be sufficient to ensure that the pennant winner is fully deserving.

As such, it is unclear that the best team in baseball wins the pennant each season. In contrast, I have come to believe that the best team in the league regularly wins the Super Bowl, the NBA championship, and the Stanley Cup[1].

Today's degree of parity is another reason why the best team in the league may not win the pennant (or, equivalently, the pennant winner may not be the best team). From a purely statistical point of view, the closer the teams are bunched together in quality, the more likely it is that the pennant winner was not really the best team in the league.[2]

In this article, I want to investigate these issues. I will construct a simulation which adjusts the degree of parity, and then determine the relationship between the degree of parity and the frequency of the best team winning the pennant. Finally, based on the simulation results and the observed degree of parity throughout baseball history, I will estimate how likely it was for the best team in the league to have won the pennant in different eras.[3]

Simulation Results

I have constructed a simple simulation of a league-season. My first set of leagues has 8 teams and plays 154-game schedules, each team playing the other teams 22 times each during a season, to simulate 1904-1961 baseball. Each team has an "innate" team winning percentage, drawn from a normal distribution with mean .500 and a fixed standard deviation reflecting the degree of parity associated with the league.

I use the formula presented by Bill James (originally developed by Dallas Adams, I think) to determine the winning pct of two pitted teams. Winning Pct (A vs B) = [A*(1-B)]/[(A*(1-B)) + ((1-A)*B)] where A is the innate winning pct of team A (i.e., vs a 500 team) and B is the innate winning pct of team B. For example, if a 600 team is pitted against a 400 team, the 600 team is likely to win 69.2%.

Using this formula, I randomly determine the outcome of every game during a 154-game season, and keep track of the overall win-loss record of each team. For each specified degree of parity in the league (i.e., the standard deviation of the normal distribution), I simulate 200,000 seasons.

The "best" team is defined to be the team with the highest innate winning percentage; ties are broken randomly if more than one team has the highest innate wpct. The pennant winner is the team with the most wins; again ties are broken randomly if more than one team winds up with the most wins in the league.

8 Teams, 154 Games

Table 1 presents the results for the 8-team simulations.

---

[1] These other sports pit team vs. team and therefore allow the spectator to see one team dominate another. The degree of domination can be reflected in a short series, even one game. Such does not appear to be the case in baseball, largely because baseball is a repeated pitting of batter vs. pitcher (not team vs. team).
[2] Parity has many benefits that probably outweigh the "downside" I am investigating here.
[3] This article is an extension to an article by Bill James in the 1989 Baseball Abstract.

The first column of table 1 specifies the degree of parity in the league (the standard deviation of the underlying normal population), given in terms of winning pct. Of course, the smaller the spread, the closer the teams are in quality and the more parity is in the league.

The second column presents the average observed spread in the teams' winning pcts. Note that the entries in the second column are greater than those in the first column. There are two explanations for this phenomenon. First, in any small sample luck will play a role, where here "luck" means normal statistical variation. Consider the first entry in the column where the observed spread is .041. In this league all teams have equal abilities (innate spread of .000). Of course, it is highly unlikely that every team in the league will go exactly 77-77. The .041 reflects the fact that some teams will win more than 77 games, and some teams less, due entirely to luck.[4]

Table 1 – 8-Team Simulation (154-game schedule)

Innate Team WPct Spread	Observed Team WPct Spread	Avg Pennant WPct	Frequency Best Team Wins Pennant (%)
.000	.041	.561	12.5
.010	.045	.563	20.9
.020	.049	.569	30.0
.030	.055	.577	38.9
.040	.063	.587	46.6
.050	.071	.598	52.9
.060	.080	.610	58.3
.070	.090	.623	63.0
.080	.100	.636	66.6
.090	.110	.649	69.8
.100	.120	.662	72.3

Second, the way I programmed the simulation will cause the observed spread to exceed the innate spread. The innate spreads essentially are valid if the opposition always has a .500 winning percentage. However, this is not the case. For example, suppose there are two teams in the league with innate winning pcts of .600 and .400. By the Adams-James formula, we would expect the observed winning pcts of these teams to be .692 and .308 when facing each other. Thus, the observed spread will be wider than the innate spread. Accordingly, the information contained in the table may be useful for others who perform baseball simulations.[5]

The third column presents the average winning pct of the pennant winners. The first row indicates that when all teams are of equal abilities, the pennant winner typically wins 86 games in a 154-game schedule (as compared to its "expected" 77). As the degree of parity decreases and the teams' abilities are more spread out, the pennant winner wins more games, ultimately winning about 102 games (out of 154) when the innate spread is .100.

The fourth column presents the percentage of times that the best team in the league turned out to win the pennant, where ties are broken randomly. The first row reflects the limiting case; here all teams are of equal abilities, so that the "best" team wins the pennant 1/8th of the time (12.5%). In this case, the outcome of each game is essentially a 50/50 coin flip, so that winning the pennant is due entirely to luck. As the teams' abilities are more spread out, the best team wins the pennant more often, ultimately winning over 72% of the pennants when the innate spread is .100.

12 Teams, 154 Games

Table 2 presents the results for a second set of simulations of leagues with 12 teams, again playing 154 game schedules, each team playing each opposing team 14 times[6]. The results are fairly similar to the previous table.

---

[4] If all teams are of equal abilities, each game is essentially a coin-flip, so that the binomial assumptions apply. The well-known formula for the standard deviation of the proportion of successes is given by SQRT[p*(1-p)/n], where SQRT denotes the square root. At p=0.5 and n=154 games, the formula gives 0.040, as verified by the entry in the table.

[5] Clearly, the strength of two factors moves inversely. The impact of luck is highest when parity is greatest, and the impact of the James-Adams "log5" formulation is highest when parity is least. By running additional simulations without the log5 formula, I can estimate how much each of the two reasons contributes to the additional spread. When the true spread is .000, the "luck" factor is the sole contributor; when the true spread is .045, luck's contribution is 60% and the log5 method's contribution is 40%; when the true spread is .095, luck's contribution is 25% and the log5 method's contribution is 75%.

[6] For simplicity, I do not model divisions or playoffs. I take the pennant winner to be the team with the most wins during the regular season. Relatedly, I did not simulate 162-game seasons, though the results should be similar.

Actual League Data

I now want to make use of the simulation results to estimate how likely the best team won the pennant throughout baseball history. As the simulations indicated, a key determining factor is the spread in the teams' winning percentages. Accordingly, I tabulated this data in different eras during the 20th century.

Table 3 presents the actual team spreads. 1904 is the first year I looked at as it was the first season with a 154-game schedule. The table generally presents the results by decade. I have split the "modern" era into 1961/2-75 as the post-expansion pre-free-agency era, and 1976-99 as the post-expansion post-free-agency era.[7]

Table 2 – 12-Team Simulation (154-game schedule)

Innate WPct Team Spread	Observed Team WPct Spread	Avg Pennant WPct	Frequency Best Team Wins Pennant (%)
.000	.041	.568	8.3
.010	.044	.571	15.3
.020	.048	.576	23.8
.030	.053	.585	32.4
.040	.060	.596	40.0
.050	.069	.608	46.7
.060	.077	.621	52.3
.070	.085	.634	57.1
.080	.095	.649	61.4
.090	.105	.663	64.8
.100	.115	.677	67.8

Recall that the degree of parity in the league is inversely related to this spread. Thus, the table indicates that the degree of parity has increased greatly over time. In the early decades of the century, the observed spread in team winning percentages was about .100 (15 games out of 154), allowing the typical pennant winner to win about 100 games (out of 154). Parity has increased until now the spread in teams' winning percentages is around .065 (10 games out of 162), allowing the typical pennant winner to win about 98 games (out of 162).

Table 3 – Team WPct by Decade

Period	Teams in League	Avg Pennant WPct[1]	Team WPct Spread	Implied Pct Best Team Wins Pennant
1904-09	8	.660	.118	72
1910-19	8	.648	.100	70
1920-29	8	.627	.091	65
1930-39	8	.640	.099	68
1940-49	8	.636	.093	67
1950-61/2	8	.626	.089	65
1961/2-75	10-12	.613	.070	49
1976-99	12-16	.608	.065	46

The final column comes from the previous simulation tables. For example, the first row indicates that the average pennant winner's winning pct in 1904-09 was .660, with a typical spread in team winning pcts of .118. Looking at our first simulation table (the results for the 8-team leagues) tells us that these data are consistent with a spread among the teams' "innate" winning pcts of about .100. And in such a league, the best team can be expected to win the pennant about 72% of the time.

Generally speaking, the table tells us that about 65-70% of the time the pennant winner was likely to be the league's best team from 1904-1961/2. A couple of round of expansions lowered this to about 50% in 1961/2-75. Free agency and a couple additional rounds of expansions further lowered this percentage to about 45% today.

[7] The table excludes shortened seasons such as 1918, 1994, 1995, and the first year following each league expansion. Also, the "pennant" winner is taken to be the team with the best regular season record in each league.

Comparison to Bill James' Study

In the 1989 Baseball Abstract, Bill James presented a study very similar to this one. James performed a simulation to investigate the impact of "luck" on a team's won-loss record. He simulated 1,000 seasons of 162 games each with 12 and 14 team leagues, and each league broken into two divisions. He found that the best team in a division only won 54.6% of the division titles, a "surprisingly low" percentage. James found that the observed spread in team's winning percentages exceeded the "true" spread, as my tables confirm. James basically had two main conclusions. First, luck plays a large role in a team's won-loss record. Second, and related, predicting division or pennant winners is a difficult task.

Due to increases in computing power, my simulations were over 200,000 seasons instead of 1,000, so we can be sure that the results are valid. In addition, my simulations are based upon a parametric distribution of team winning percentages that can easily be varied.[8] My results are consistent with James's article. I too find that the best team only wins the pennant about half the time in today's baseball. My tables of results according to the degree of parity in the league may be useful to other researchers as well. Finally, I tied my simulation results to the historical degree of parity to estimate the likelihood that the best team in a league won the pennant throughout baseball history.

Concluding Remarks

Parity has brought with it many important benefits. Teams no longer languish in the second division for extended periods like they used to in the "good old days". Today, due to the cumulative effects of the draft, free agency, expansion, and revenue sharing, most teams can expect to compete for their division title in any given season. Recall that from 1978 to 1987, 10 different teams won the World Series, and in the past twenty seasons there have been 15 different World Champs. Fans stay interested over time and deeper into seasons.

However, from a purely statistical point of view, increased parity has the negative consequence of allowing inferior teams to make the playoffs and ultimately win the pennant. Using simulation analysis and a historical review of teams' winning percentages, I estimate that the best team in the league was likely to win the pennant more than 70% at the beginning of the 20th century. The increased degree of parity exhibited in today's era has driven this percentage below 50%. Today, the team with the best record in the league is less than 50% likely to be the best team in the league.

Rob Wood is a management consultant in California. Rob Wood, 2101 California St. #224, Mountain View, CA, 94040-1686, rob.wood@us.pwcglobal.com. ♦

[8] I sample from a normal distribution.