Right Questions, Wrong Answers -- A Review of "The Wages of Wins"

Phil Birnbaum

From the May, 2006 issue of "By the Numbers"

Introduction

A good description of "The Wages of Wins" [TWOW] is Alan Schwarz's cover quote: "Freakonomics meets ESPN." Authors David J. Berri, Martin B. Schmidt, and Stacey L. Brook, three academic economists, analyze data from three sports to, in the words of the subtitle, "[take] measure of the many myths in modern sport."

While the subject area is ostensibly economics, only a couple of the chapters deal with traditional sports economics issues. The majority of the text attempts to find performance measures for basketball and football; the subject matter could easily be described as sabermetrics of several sports. Indeed, the authors are serious followers of basketball -- almost half the book is devoted to analysis of that sport.

To its credit, the authors describe the studies and results that lead to their conclusions, unlike "Freakonomics," which discuss the implications of the findings rather than the logic that led to those findings. Some of their regression findings are presented (in suitable simplified form), so readers can connect the studies' results to the authors' conclusions.

The book has garnered good reviews and blurbs, from Alan Schwarz to (famed economist) Deirdre McCloskey, to a long New Yorker review by Malcolm Gladwell. The book is indeed a decent read.

However, as for the details of the authors' studies, in my judgment, many of the authors' conclusions are incorrect.

Wins and Payroll

The chapter entitled "Can You Buy the Fan's Love?", for instance, discusses the effects of MLB team payroll versus performance. The authors regress payroll on wins, and find an r-squared of .176. From this, they argue that "payroll and wins are not strongly linked," because the explanatory power of wins is only 18%.

But an r-squared of .176 *is* a strong relationship between payroll and wins! The statement "payroll explains 18% of wins" is true only in a narrow, mathematical sense -- that the reduction of sums of squares from the regression line after adjusting for salaries is 18% of the total sums of squares from the mean before the adjustment. It doesn't answer the relevant question, which is: what is the relationship between payroll and wins in the baseball context?

For that question, the important number is not the r-squared of .176, but its square root, the correlation coefficient. The square root of .176 is about .42. This means that for every additional standard deviation in salary a team spends, it will improve .42 of a standard deviation in wins. Roughly speaking, 42% of a team's spending will show up in the win column.

That's pretty large, considering how much luck there is in a team's record. A superstar is worth about five wins above an average player. The SD for a team of known talent is six wins. That means that almost half the time, luck is a bigger factor than, say, Derek Jeter. Under these conditions, a correlation of .42 is a fairly large factor.

The authors note that $5 million in payroll buys an additional win. "That's all the bang a team would get for their buck," they write. " … a team would have to add several $10 million players before they could expect to see any real progress in the standings."

But, of course, teams *are* spending that much. In 2005, the Yankees spent about $140 million above the median. The difference, according to the authors' study, should be about 28 wins above average (for an expected 109-53 record). The Red Sox, at $123 million, would start with a 14 win advantage. And the Devil Rays, with a payroll of only $30 million, would have a 7-game disadvantage.

So if the Yankees start with an expectation 35 wins above Tampa Bay, and the Red Sox start 21 games ahead, doesn't that imply that salary is important?

The authors, again, say no, because of the 18% figure.

This might be a nitpick if the authors' argument was a minor one. But it's the primary thesis of the chapter, and the inspiration for the title of the book. The authors start the chapter by arguing , reasonably, that looking at post-season victories is not a good way to check the effects of salaries. They argue, again reasonably, that MLB's "Blue Ribbon Panel" looked only at 1995-1999, the period in which salary had the largest effect. But when it comes to data to actually determine the relationship in the way the authors consider most appropriate, this is the only study they use. And they draw a conclusion that contradicts what the data actually show.

Competitive Balance

There is a fear that, if revenue inequality among teams continues to increase, some teams will buy up all the free agents, other teams won't be able to afford any, and competitive balance will continue to worsen, perhaps fatally to the game.

In Chapter 4, the authors argue against this eventuality. First, they write, studies have shown that attendence rises the more uncertain the outcome of the game. If the Yankees turn into a team of Babe Ruths, and the Royals turn into a team of Danny Ainges, a Yankee victory is almost preordained, and fans won't show up to the game, even in New York. So the Yankees have a vested interest in leaving enough quality players for other teams.

In addition, adding more and more superstars has less and less effect on a team's win totals, and therefore revenue. As a team gets better, there are fewer and fewer losses for it to turn into wins. Derek Jeter might be worth five wins to an average team, but only, say, three wins to the 1998 Yankees. Again, this puts a natural limit on how much a wealthy team will be willing to spend on players, and, by extension, on competitive imbalance.

Not only is there a natural limit to competitive balance, but revenue sharing won't even help. One of the most famous results in economics is the Coase Theorem, which says that resources (players) wind up going to the firms (teams) to whom they have the most value, regardless of who actually owns them at any given time.

The Wages of Wins: Taking Measure of the Many Myths in Modern Sport

By David J. Berri, Martin B. Schmidt, and Stacey L. Brook

Stanford University Press, 304 pages, $29.95 (US), ISBN 0465005969

That is, suppose Derek Jeter is worth $10 million in revenue (increased attendance, TV viewership, playoff chances) to the Yankees, but only $5 million to the Pirates. The Coase Theorem says that even if you force the Yankees to share revenue with the Pirates, and even if you allow the Pirates to draft Jeter in the first place, he will be sold to the Yankees (or traded for cheaper players) simply because he's worth more in New York.[1] This is good stuff and well explained.

Finally, the authors show measures of competitive balance (as measured by W-L records) for fifteen different sports leagues. It turns out that soccer has the most balance, followed by football, hockey, baseball, and, most unbalanced, basketball.

The authors argue that this is the result of the population from which players are taken. Hundreds of millions of people in the world play soccer, so there is an abundance of talent. But basketball is limited mostly to people who are very tall, and there are a lot fewer of those. So, with tall basetball players in short supply, the leagues are filled with inferior players, which allow the best players to dominate.

It sounds plausible, but the problem is that it's the rules of the particular sport that determine competitive balance. A simple thought experiment can show why this is true. Suppose that instead of basketball games being 48 minutes long, they were 480 minutes long. With ten times the opportunity for luck to even out, the better team is much more likely to win the game – perhaps it becomes a .900 team instead of a .600 team. Or suppose the game is shortened to 4.8 minutes. In a game that short, even a markedly inferior team could win; perhaps it becomes a .450 team instead of a .300 team.

Competitive balance is partly a consequence of the rules of the game. The more opportunity for luck to even out, the more likely the better team is to win, and the more unbalanced the sport looks.

[1] One implication, though, which the authors do not address, is that a payroll tax *would* increase competitive balance. If the Yankees have to pay a $10 million tax on Jeter's $4 million salary, but the Pirates don't, he becomes unprofitable to the Yankees ($14 million cost, $10 million benefit) but remains profitable to Pittsburgh ($4 million cost, $5 million benefit).

Let's compare baseball and basketball. I'll oversimplify a bit to make the comparison easier, but the argument will stand even if the details are made more realistic.

- In basketball, each team has 100 ball possessions in which to score; each possession, they score about 50% of the time. In baseball, each team has about 40 plate appearances in which to put men on base; in each plate appearance, they get on base about 40% of the time.

  So basketball has 2.5 times as many chances for the better team to assert its superiority. Since the standard deviation of scoring rate is proportional to the square root of opportunities, we can say that basketball has over 50% more imbalance than baseball in this respect.

- In basketball, the team that scores most wins. In baseball, the team that gets on base more doesn't necessarily win -- it depends *how* it gets on base, and whether those successes are bunched into a relatively few innings.

  So, in basketball, the team with the better success rate wins. In baseball, there's a big "luck" factor (assuming that clutch hitting ability does not dominate) that allows the weaker team to beat the stronger team despite being outperformed in the basic success rate. Again, in this regard, basketball has much more imbalance than baseball.

- In basketball, five players play most of the game. The average player is 20% of his team's performance. In baseball, nine hitters play most of the game. The average player is 11% of his team's offense. (And even a workhorse starting pitcher is only 15% or so of his team's innings.)

  So in basketball, the average player has much more impact on the outcome than in baseball. That allows the superstars to play more, and the average players less, which again means more imbalance in basketball.

- In basketball, a superstar might take 40% of his team's shots. In baseball, every player must bat in turn, so even the leadoff hitter will have no more than about 12% of his team's plate appearances.

  Again, this means more imbalance in basketball, because the superstars can dominate.

Add up these factors, and it becomes evident why basketball has less competitive balance than baseball; the game is structured so that the team with the best players is much more likely to win. These are four powerful theoretical reasons why records should be more extreme in basketball even before considering demographics.[2]

Is there any empirical evidence that can be examined on the question? There is: home field advantage.

In baseball, the home team has a winning percentage of about .540 – 40 points above normal. In basketball, the home team has a winning percentage of around .625 – three times as high.

In 2002-2003, the .488 Seattle SuperSonics were .610 at home (the equivalent of 99-63), but only .366 on the road (the equivalent of 59-103). This can't be explained by the demographics of height.

But it *can* be explained by game structure. Home field makes one team better and the other team worse. In basketball, the game is structured so that the better team wins with high probability. In baseball, the game is structured so that the better team wins with lower probability.

Another way of phrasing the difference is that the structure of basketball gives the illusion that basketball leagues are more unbalanced in terms of talent. They may be; they may not be. But they certainly are more unbalanced in terms of game results.

It's still possible that the authors' "short supply of tall people" theory makes some contribution, but I'm not convinced. It is indeed true that "you can't teach height," which means that the supply of basketball players is limited to a small subset of the population. But, in baseball, "you can't teach sight" either -- if Ted Williams really had 20/10 eyesight, and if extremely good vision is as important in baseball as height is in basketball, you have an analogous situation for baseball. There are probably many important attributes in any sport that are intrinsic and can't be taught, some of which we're not even aware of. (Wayne Gretzky was able to "see" the game in slow motion.) In the absence of any argument that height in basketball is more important than similar attributes in other sports, I remain agnostic.

---

[2] After the first draft of this article was written but before going to press, another review of this book, by Roland Beech, made this same point. That review can be found at http://www.bepress.com/jqas/vol2/iss3/5/ .

Basketball

This being a baseball publication, I will summarize the extensive basketball portion of the book only briefly.

In 2003, Dean Oliver's "Basketball on Paper" was published. Of the basketball books I've seen, it's the closest in spirit to the Bill James Baseball Abstracts, and many of Oliver's insights are reminiscent of James. For instance, just as Bill James noted that the out (and not the at-bat) is the currency of a baseball offense, Oliver notes that for basketball, it's points per possession, and not points per game, that's important. That's because every game has a different pace and a different number of possessions, so a team that scores 100 points per game in 100 opportunities per game is a better offense than a team scoring 102 points per game in 105 opportunites per game.

Oliver argues that players should be evaluated per posession. A player would be credited with his team's possession when he ends it -- by shooting, taking free throws, or turning the ball over. A player is credited with the points he is responsible for, but assists are divided between the scorer and the passer.

It's from this base of knowledge that "The Wages of Wins" begins.

First, the authors note that since a possession, on average, scores one point, then a turnover costs one point (being the forfeiture of a possession). A field goal made turns a one-point potential into a two point score, so is also worth one point above average. And a successful 3-point shot is obviously worth two points. The authors list a chart of almost everything that can happen during a possesion, and how many points it's worth. They note that every 30 points equals one win (like 10 runs in baseball), and so they divide by 30 to get a win value for each event in terms of wins. Then, because guards and frontcourt players have different per-game averages, they adjust for position.

And for defense, they take the team's defensive bottom line, and allot it to players based on minutes on the court.

But then, they make what I think is a critical error. In converting player wins into a rate, they divide by minutes played, instead of by possessions used.

This is a problem for several reasons. First, consider two players who always play together; they have exactly identical results per shot, but one takes twice as many shots as the other. Using minutes, one will look twice as good as the other. Second, teams who play a faster-paced game and get more possessions than average will have their players look better than equally-talented players who play on slower teams. Finally, you'll get wrong results when composing hypothetical teams. If the 1995-96 Michael Jordan scored .386 wins per game (48 minutes), does that mean five Jordans would have scored five times that? No -- because if Jordan takes 40% of his team's possessions, five Jordans would need to take 200% of the team's possessions, which isn't possible.

These issues can seriously change conclusions and rankings. Some teams are almost ten percent faster than others (meaning 10% more points can be scored by players on those teams). Further, some players take almost 30% of their team's possessions – one-and-a-half times as much as average. A player could rate some 65% higher than another similar player for these reasons alone.

This flaw, in my opinion, renders three chapters worth of results unreliable, at least those results based on the rate stat.

Clutch and Consistency

In an attempt to find whether some basketball players improve in the clutch, the authors compare players' performances in the regular season to those in the playoffs. They find that performance generally drops. Could that be simply because, in the playoffs, their opponents are limited to the league's better teams? Astoundingly, in eight pages of analysis, the authors don't mention the possibility even once!

Finally, the authors turn to players' consistency. Using a percentile grading system, they find that only 12% of basketball players moved more than two grades (up or down) between seasons. In MLB, 28% moved up or down two grades. But for NFL quarterbacks, the figure was 39%. The authors argue that quarterbacks are "consistently inconsistent," and the difference is caused by team factors such as the quality of the offensive line. That's probably a substantial part of the answer, but I suspect most of the difference is just luck. Quarterbacks have fewer opportunities than basketball players (and about the same as baseball players) – and the standard deviation of outcomes is very high.

Regardless of the reasons, I agree with the authors' conclusions that overreliance on QB statistics is a mistake.

Complexity

Some of the most difficult and interesting problems in the analysis of free-flowing sports is how the players affect each other in ways that are hard to measure.

For instance, suppose player A scores more assists than player B on another team. Is that because A is better at hitting his man near the basket, or is it because his teammates are better at getting open?

If player C is more successful than player D, is it because C is better at shooting, or because the opposing defense is concentrating on covering Michael Jordan, giving C more room to get to the basket?

If player E has a poor shooting percentage, is that because he's not accurate, or because his team often gets him the ball with no time left on the 24-second clock, forcing him to take desperation shots?

All these factors influence a player's stats, but the book doesn't study any. And that's fine -- with the limited statistical record the authors are limited to working with, it probably isn't possible. But what's frustrating is that the authors don't even mention them. Having created and run their model, they proceed as if the problem has been solved.[3]

The authors are very critical of Allen Iverson, arguing that his productivity ranks far below his reputation and traditional scoring statistics. And despite the imperfections of the measures used, the numbers are fairly convincing that the authors are correct. However, given that the numbers highly contradict conventional wisdom, might there be factors the authors didn't consider?

It took me only a few minutes of web searching to find an article suggesting that Iverson, who takes many of his team's shots, saves his teammates from having to take difficult shots by doing so himself. That might explain the disconnect between his high scoring and his low overall rating. It would also suggest that he's not as bad as TWOW suggests, as he's taking a hit to his personal stats to the benefit of his teammates' stats (and hopefully, to the benefit of the team too, if he's the player with the best chance of sinking the hard shots).

Now, I don't know if this is true. But it might be, and the authors don't give it a thought.

Football

TWOW's football chapter can be summarized by one regression result: to evaluate a quarterback's productivity, (1) take his total yards passed plus rushed; (2) subtract 3 for every play; and (3) subtract 50 for every turnover. The result is followed by several pages of player charts and commentary.

I'm sure the formula accurately reflects the regression result. The problem with this method is that it doesn't take the situation into account.

A two-yard rush is worth negative one point by this system, but on third-and-one, it's exactly what's called for, and so an unqualified success. And three consecutive passes of four yards each are worth exactly as much, in real life, as a single 12-yard pass, but the system values them differently. Quarterbacks with different teams who can run more plays (but be equally successful in gaining enough yards to maintain possession) will be underrated by this formula; quarterbacks who lose the ball equally as often after an equal number of yards, but in fewer plays, will be overrated.

Again, I'm not sure how accurate or inaccurate the system is given these difficulties. However, the authors seem unaware of these issues, arguing that it's "just about everything one could want in a performance measure." But a more sophisticated approach was given by Carroll, Palmer, and Thorn in "The Hidden Game of Football" in 1989. That book does not appear in TWOW's bibliography.

Overall

There is a common pattern the authors use throughout the book – run a regression, explain the findings, assume the problem is now solved, then dismiss conventional wisdom because it doesn't use regression. On page 7, the authors express dismay at the "laugh test" – the tendency to dismiss analytical findings if they contradict conventional wisdom -- correctly pointing out that research trumps intuition. But the authors go too far the other way. With equal lack of justification, they unthinkingly reject any non-statistical opinion that contradicts their results. Which I think is why they're off the mark so often: they fail to consider that their analysis may be incomplete, or may not

---

[3] Actually, the authors do consider whether a player can make his teammates better -- but by "better," they mean "better in their total value statistics." And, with Michael Jordan taking a high percentage of his team's opportunities, other players' stats are not going to improve when MJ shows up.

completely capture what they're trying to measure. They give opposing views no benefit of the doubt, and their own views get no doubt at all.

And so readers with a decent knowledge of sabermetrics will find this book frustrating -- there isn't much new, the authors are inappropriately immodest, and many of the results, I think, are just plain wrong. One particular frustration is that the authors seem unaware of previous research. On page 41, when claiming that money doesn't buy MLB wins, they suggest that maybe GMs don't know that players decline after age 28. Bill James' study on aging dates back to 1982, and players peaking in their late-20s has become conventional wisdom since then. Not only are the authors unaware of this (their reference is an academic working paper from 2005!), but they blithely assume that general managers know little about the product they're putting on the field.

But having said all that, the book may still be worth a look, if you can get the gnashing of teeth every page or two. The authors write clearly, and they raise interesting questions. If you're looking for the answers, this may not be the place -- but there are many studies that suggest themselves out of the issues the authors raise.

Phil Birnbaum, birnbaum@sympatico.ca ♦