Article from

Predictive Analytics and Futurism

July 2016
Issue 13

The Actuarial Road Not Taken: Jeff Sagarin's Sports Ratings

By Anders Larson

To close out the 2015-16 college football season, the Clemson Tigers were set to play the Alabama Crimson Tide. Like many matchups in college football's postseason, the two teams had not played each other during the regular season. In fact, they had not played each other since 2008.

So who should be expected to win? Well, one place to look would be the point spread. The point spread for each game is devised by gambling organizations, and it is designed to handicap the game such that each team should have roughly an equal chance to win, after adding in the point spread.[1] For instance, the point spread between Clemson and Alabama was seven points in favor of Alabama at most sports books, meaning gamblers betting on Alabama would need Alabama to win by at least seven points to win the bet.

Point spreads have been around for years, and ever since their inception, sports gamblers and casual fans alike have looked for ways to outsmart the oddsmakers. Back in the early 1970s, one of those fans was Jeff Sagarin, a recent graduate of the Massachusetts Institute of Technology. At the time, Sagarin was considering a career as an actuary. He passed three actuarial exams in the late 1960s and early 1970s and worked at New York Life as an actuarial trainee for a brief period.

Many gamblers would look to superstition or misguided "statistics" (hey, the Yankees have won their last seven Tuesday games against a right-handed pitcher—they're a lock![2]). Sagarin took a more analytical approach: He decided to devise a rating system that would help predict both the outcome and the margin of victory if two teams played each other. His ratings differed from the traditional poll rankings, which often came down to a subjective opinion of which teams were most "deserving."

"I never even thought about the 'reward' thing," Sagarin said. "I wanted to predict games as accurately as possible. I wanted to see if I could be as accurate as the point spreads in the New York Post."

Sagarin's system was data-driven, taking into account scores from games across the country (and other variables, such as home-field advantage). Aggregating all the scores was critical, but at the time, simply collecting those scores was often the hardest part of the process. As actuaries, we understand the challenge of aggregating and cleaning data and how critical it is to our work product. But most of us who entered the industry in the past 20 years probably haven't had to go to the same lengths that Sagarin did back in the 1970s. On a typical day, Sagarin, living in Boston, would drive to the office of the Boston Globe and cut out its wire with each day's basketball scores. But since the Globe's wire was often incomplete, Sagarin had to resort to plan B.

"So sometimes what he'd have to do was, say there was a small school, say Ball State in Indiana," said Larry Isaacs, a long-time friend of Sagarin's. "What he'd do is call telephone information in Muncie, Indiana, and get the operator on the line, and he'd sweet talk the operator because he was pretty charming. And he'd say, 'By the way, was there a basketball game last night?' And he'd say, 'Can you tell me who won that game?' That's how we'd get the scores. It was really low-tech in those days."

After finding he was having some success, Sagarin caught on with some magazines and newspapers, including Pro Football Weekly and the Boston Globe. He ultimately decided to forego actuarial science and make a full-time career out of his sports rating systems. His big break came in 1985, when his ratings started appearing in USA Today, where they still appear for several sports. Sagarin's ratings, which have been around far longer than most other rating systems currently on the market, consistently have among the most accurate predictions. ThePredictionTracker.com evaluates nearly 70 college football rating systems.[3]

Sagarin's primary rating system predicted winners more accurately than all other comparable computer-based systems[4] that were ranked in each year from 2013 to 2015. His ratings correctly predicted approximately 76 percent of games, which was better than the opening betting lines (although the midweek and updated betting lines were slightly more accurate). Keep in mind that Sagarin's ratings do not account for some information that oddsmakers use to set the betting lines, such as recent injury reports, player suspensions or weather.

His ratings were also used in the Bowl Championship Series (BCS), which determined the college football national championship game participants from 1998 to 2013. Sagarin said his involvement with the BCS was a blessing and a curse. The NCAA wanted to use his rating system to help pick the teams for the national championship but with a caveat.

"The NCAA told me, 'We know you need to use scores,'" Sagarin said. "'We're all coaches, we know the score tells you a lot. But as the NCAA, we can't officially have a rating system that uses the scores. Our official system can only take into ac-

count winning and losing.' They initially didn't even take into account home and away games!"

As a result, Sagarin came up with a system that ignored the actual scores of games, focusing almost entirely on wins and losses. He referred to it as the "Elo" system because of its similarity to the chess ratings developed by Arpad Elo in 1950.

Certainly at this point, actuaries from all disciplines can likely relate to Sagarin's dilemma. With almost any predictive modeling technique, using more information should generally yield more accurate results, if the modeling is done responsibly (for instance, by avoiding overfitting). However, there are often reasons certain variables need to be excluded. For instance, under President Obama's health care law, the Patient Protection and Affordable Care Act, health insurers on the individual exchange are only allowed to modify the base premium for an individual based on age, smoking status and geographic area.

Another commonality between Sagarin's current line of work and actuarial science is the blend of statistical competence and subject matter expertise. Sagarin has a variety of rating systems, but he began with a simple exponential smoothing system. Exponential smoothing is a technique for smoothing time series data. It uses all the historic information available but makes recent observations worth more than older ones. The actual balance of credibility between the recent observations and the older ones needs to be tuned for each forecast. It is possible to tune this balance by formulating a data-generating process and then

minimizing errors, but just as often it is tuned by subjective domain knowledge and common sense.

In the case of Sagarin's rating system, the technique was used to modify the rating of a team on the basis of recent results, but it still required the smoothing factor to be chosen by the modeler to determine how much value to place on the recent results. Setting this factor appropriately is where Sagarin's intuition and knowledge of the game came into play.

"Let's say two teams play and you had them rated equally and one wins by two touchdowns. How much do you change it?" Sagarin said. "One person who knows nothing about sports would say, how about you put it to seven? Well, only a moron would do that. Teams don't change like that. You don't want to change the ratings like that. You only want to move it by a couple points."

This concept is not entirely different from the concept of actuarial credibility. If a health insurer offers coverage to a new large group, the premium is often set based on a "manual rate" based on the available information about the group, such as the demographic mix. However, at the start of the next year, the insurer likely would charge a premium that reflected a blend of the manual rate and the group's actual observed claims costs. The amount of weight given to the group's actual experience is referred to as the "credibility" of the group. Although there are commonly used formulas to estimate the credibility based on the size of the group and other factors, actuaries also have to rely on their own judgment.

Starting with an appropriate manual rate is critical to successful pricing in the insurance industry. A similar concept applies with sports ratings. Although there are some rating systems that are independent of the pre-season ratings, such as a Simple Rating System,[5] they can often produce unrealistic results early in the season. Other systems, such as those that rely on Bayesian concepts, rely heavily on the starting values (referred to as a priori estimates in a Bayesian framework).

"Bayesian systems will get better predictive results than going with pure unbiased results because [the pure unbiased results] are based only on the games early in the season," Sagarin said. "By mid-season, sort of 'All roads lead to Rome,' and all of the prediction systems are pretty similar."

Sagarin said the starting values are crucial to success. His starting values are based on a time-series analysis of each team's rating history.

"If you have good starting ratings, you'll have good ratings all year long," Sagarin said. "If you start off with Ohio State as the worst team in the country and Columbia as the best team, you're going to have problems."

Isaacs, a fellow of the Society of Actuaries who works with the IRS, said he "absolutely" sees similarities between Sagarin's work and the work we do as actuaries.

"After you've been doing this a while, you just have a sense that something's right or wrong," Isaacs said. "You just know. It's the same thing with his rating systems. You just have a sixth sense about what makes sense and what doesn't. I do that all the time. Someone will show me some numbers, and I'll say 'Something's not clicking, something just doesn't feel right.' I think there's a lot of carry-over to what Jeff does. He looks at the scores, but he looks at home and away, and all sorts of other things. To do what he does, he's got to be doing more than just plugging in scores."

I made a major career transition myself after working as a sportswriter in Columbus, Ohio, for more than three years. In 2009, I decided to begin taking actuarial exams and eventually was fortunate enough to secure a job at Milliman. At the time, I imagined the intersection of sports and actuarial science was virtually nonexistent. Seven years later, I can see I was wrong. The principles of actuarial science extend far beyond the insurance industry, as long as you know where to look. ∎

Anders Larson, FSA, MAAA, is at Milliman in Indianapolis. He can be reached at *anders. larson@milliman.com*.

ENDNOTES

[1] To be precise, the point spreads are actually devised to entice an equal amount of money to be wagered on each team so the gambling organization is guaranteed to make a profit, after accounting for the fees they charge gamblers.

[2] Ironically, given the same evidence, many gamblers would probably claim the Yankees couldn't possibly win eight consecutive Tuesday games against right-handed pitchers, and therefore they were "due" to lose.

[3] *http://www.thepredictiontracker.com/predncaa.html.*

[4] The term "comparable computer-based systems" excludes the actual betting lines and other systems that incorporate the betting line.

[5] Doug Norris, "Simple Rating Systems: Entry-Level Sports Forecasting." Forecasting and Futurism (July 2015), *https://www.soa.org/library/newsletters/forecasting-futurism/2015/july/ffn-2015-iss11-norris-2.aspx.*