How Our NFL Predictions Work

See our latest predictions

---

references

Pro-Football-Reference.com

Autocorrelation / Elo rating / Monte Carlo simulations / Regression to the mean

the details

FiveThirtyEight has an admitted fondness for the Elo rating — a simple system that judges teams or players based on head-to-head results — and we've used it to rate competitors in basketball, baseball, tennis and various other sports over the years. The sport we cut our teeth on, though, was professional football. Way back in 2014, we developed our NFL Elo ratings to forecast the outcome of every game. The nuts and bolts of that system are described below.

Game predictions

In essence, Elo assigns every team a power rating (the NFL average is around 1500). Those ratings are then used to generate win probabilities for games, based on the difference in quality between the two teams involved, plus the location of the matchup. After the game, each team's rating changes based on the result, in relation to how unexpected the outcome was and the winning margin. This process is repeated for every game, from kickoff in September until the Super Bowl.

For any game between two teams (A and B) with certain pregame Elo ratings, the odds of Team A winning are:

$$Pr(A) = \frac{1}{10^{\frac{-EloDiff}{400}} + 1}$$

ELODIFF is Team A's rating minus Team B's rating, plus or minus a home-field adjustment of 65 points, depending on who was at home. (There is no

home-field adjustment for neutral-site games such as the Super Bowl[1] or the NFL's International Series.) Fun fact: If you want to compare Elo's predictions with point spreads like the Vegas line, you can also divide ELODIFF by 25 to get the spread for the game.

Once the game is over, the pregame ratings are adjusted up (for the winning team) and down (for the loser). We do this using a combination of factors:

- **The K-factor.** All Elo systems come with a special multiplier called *K* that regulates how quickly the ratings change in response to new information. A high K-factor tells Elo to be very sensitive to recent results, causing the ratings to jump around a lot based on each game's outcome; a low K-factor makes Elo slow to change its opinion about teams, since every game carries comparatively little weight. In our NFL research, we found that the ideal K-factor for predicting future games is 20 — large enough that new results carry weight, but not so large that the ratings bounce around each week.

- **The forecast delta.** This is the difference between the binary result of the game (1 for a win, 0 for a loss, 0.5 for a tie) and the pregame win probability as predicted by Elo. Since Elo is fundamentally a system that adjusts its prior assumptions based on new information, the larger the gap between what actually happened and what it had predicted going into a game, the more it shifts each team's pregame rating in response. Truly shocking outcomes are like a wake-up call for Elo: They indicate that its pregame expectations were probably quite wrong and thus in need of serious updating.

- **The margin-of-victory multiplier.** The two factors above would be sufficient if we were judging teams based only on wins and losses (and, yes, Donovan McNabb, sometimes ties). But we also want to be able to take into account *how* a team won — whether they dominated their opponents or simply squeaked past them. To that end, we created a multiplier that gives teams (ever-diminishing) credit for blowout wins by taking the natural logarithm of their point differential plus 1 point.

$$MovMultiplier = \ln{(WinnerPointDiff + 1)}$$
$$\times \frac{2.2}{WinnerEloDiff \times 0.001 + 2.2}$$

This factor also carries an additional adjustment for autocorrelation, which is the bane of all Elo systems that try to adjust for scoring margin. Technically speaking, autocorrelation is the tendency of a time series to be correlated with its past and future values. In football terms, that means the

Elo ratings of good teams run the risk of being inflated because favorites not only win more often, but they also tend to put up larger margins in their wins than underdogs do in theirs. Since Elo gives more credit for larger wins, this means that top-rated teams could see their ratings swell disproportionately over time without an adjustment. To combat this, we scale down the margin-of-victory multiplier for teams that were bigger favorites going into the game.[2]

Multiply all of those factors together, and you have the total number of Elo points that should shift from the loser to the winner in a given game. (Elo is a closed system where every point gained by one team is a point lost by another.) Put another way: A team's postgame Elo is simply its pregame Elo plus or minus the Elo shift implied by the game's result — and in turn, that postgame Elo becomes the pregame Elo for a team's next matchup. Circle of life.

Elo does have its limitations, however. It doesn't know about trades or injuries that happen midseason, so it can't adjust its ratings in real time for the absence of an important player (such as a starting quarterback). Over time, it will theoretically detect such a change when a team's performance drops because of the injury, but Elo is always playing catch-up in that department. Normally, any time you see a major disparity between Elo's predicted spread and the Vegas line for a game, it will be because Elo has no means of adjusting for key changes to a roster and the bookmakers do.

Pregame and preseason ratings

So that's how Elo works at the game-by-game level. But where do teams' pregame ratings come from, anyway?

At the start of each season, every existing team carries its Elo rating over from the end of the previous season, except that it is reverted one-third of the way toward a mean of 1505. That is our way of hedging for the offseason's carousel of draft picks, free agency, trades and coaching changes. We don't currently have any way to adjust for a team's actual offseason moves, but a heavy dose of regression to the mean is the next-best thing, since the NFL has built-in mechanisms (like the salary cap) that promote parity, dragging bad teams upward and knocking good ones down a peg or two.

Note that I mentioned "existing" teams. Expansion teams have their own set of rules. For newly founded clubs in the modern era, we assign them a rating of 1300 — which is effectively the Elo level at which NFL expansion teams have played since the 1970 AFL merger. We also assigned that number to new AFL

teams in 1960, letting the ratings play out from scratch as the AFL operated in parallel with the NFL. When the AFL's teams merged into the NFL, they retained the ratings they'd built up while playing separately.

For new teams in the early days of the NFL, things are a little more complicated. When the NFL began in 1920 as the "American Professional Football Association" (they renamed it "National Football League" in 1922), it was a hodgepodge of independent pro teams from existing leagues and opponents that in some cases were not even APFA members. For teams that had not previously played in a pro league, we assigned them a 1300 rating; for existing teams, we mixed that 1300 mark with a rating that gave them credit for the number of years they'd logged since first being founded as a pro team.

$$InitRating = 1300 \times \frac{2}{3}^{YrsSince1stSeason} + 1505$$
$$\times \left(1 - \frac{2}{3}\right)^{YrsSince1stSeason}$$

This adjustment applied to 28 franchises during the 1920s, plus the Detroit Lions (who joined the NFL in 1930 after being founded as a pro team in 1929) and the Cleveland Rams (who joined in 1937 after playing a season in the second AFL). No team has required this exact adjustment since, although we also use a version of it for historical teams that discontinued operations for a period of time.

Not that there haven't been plenty of other odd situations to account for. During World War II, the Chicago Cardinals and Pittsburgh Steelers briefly merged into a common team that was known as "Card-Pitt," and before that, the Steelers had merged with the Philadelphia Eagles to create the delightfully monikered "Steagles." In those cases, we took the average of the two teams' ratings from the end of the previous season and performed our year-to-year mean reversion on that number to generate a preseason Elo rating. After the mash-up ended and the teams were re-divided, the Steelers and Cardinals (or Eagles) received the same mean-reverted preseason rating implied by their combined performance the season before.

And I would be remiss if I didn't mention the Cleveland Browns and Baltimore Ravens. Technically, the NFL considers the current Browns to be a continuation of the franchise that began under Paul Brown in the mid-1940s. But that team's roster was essentially transferred to the Ravens for their inaugural season in 1996, while the "New Browns" were stocked through an expansion draft in 1999. Because of this, we decided the 1996 Ravens'

preseason Elo should be the 1995 Browns' end-of-year Elo, with the cross-season mean-reversion technique applied, and that the 1999 Browns' initial Elo should be 1300, the same as any other expansion team.

Season simulations

Now that we know where a team's initial ratings for a season come from and how those ratings update as the schedule plays out, the final piece of our Elo puzzle is how all of that fits in with our NFL interactive graphic, which predicts the entire season.

At any point in the season, the interactive lists each team's up-to-date Elo rating (as well as how that rating has changed over the past week), plus the team's expected full-season record and its odds of winning its division, making the playoffs and even winning the Super Bowl. This is all based on a set of simulations that play out the rest of the schedule using Elo to predict each game.

Specifically, we simulate the remainder of the season 100,000 times using the Monte Carlo method, tracking how often each simulated universe yields a given outcome for each team. It's important to note that we run these simulations "hot" — that is, a team's Elo rating is not set in stone throughout the simulation but changes after each simulated game based on its result, which is then used to simulate the next game, and so forth. This allows us to better capture the possible variation in how a team's season can play out, realistically modeling the hot and cold streaks that a team can go on over the course of a season.

Late in the season, you will find that the interactive allows you to experiment with different postseason contingencies based on who you have selected to win a given game. This is done by drilling down to just the simulated universes in which the outcomes you chose happened and seeing how those universes ultimately played out. It's a handy way of seeing exactly what your favorite team needs to get a favorable playoff scenario or just to study the ripple effects each game may have on the rest of the league.

The complete history of the NFL

In conjunction with our Elo interactive, we also have a separate dashboard showing how every team's Elo rating has risen or fallen throughout history. These charts will help you track when your team was at its best — or worst — along with its ebbs and flows in performance over time. The data in the charts goes back to 1920 (when applicable) and is updated with every game of the current season.

version history

1.1

Ratings are extended back to 1920, with a new rating procedure for expansion teams and other special cases. Seasonal mean-reversion is set to 1505, not 1500.

Sept. 10, 2015

1.0

Elo ratings are introduced for the current season; underlying historical numbers go back to 1970.

Sept. 4, 2014

related articles

The Complete History Of The NFL

May 1, 2018

Introducing NFL Elo Ratings

Sept. 4, 2014

The Best NFL Teams Of All Time, According To Elo

Sept. 18, 2015

Footnotes

1. Unless a team somehow makes the Super Bowl in its host year.

2. Special note: In the case of a tie, the multiplier becomes 1.525, or 2.2 times the natural log of 2 (which, based on the formula above, effectively assumes the absolute margin of victory in any game must be at least 1).