Chapter Five

Elo's System

Árpád Élö (1903–1992) was a Hungarian-born physics pro-
fessor at Marquette University in Milwaukee, Wisconsin. In
addition, he was an avid (and excellent) chess player, and
this led him to create an effective method to rate and rank
chess players. His system was approved by the United States
Chess Federation (USCF) in 1960, and by Fédération Inter-
nationale des Échecs (the World Chess Federation, or FIDE)
in 1970. Elo's idea eventually became popular outside of the
chess world, and it has been modified, extended, and adapted
to rate other sports and competitive situations. The premise
that Elo used was that each chess player's performance is a
normally distributed random variable X whose mean μ can
change only slowly with time. In other words, while a player

Arpad Elo

might perform somewhat better or worse from one game to the next, μ is essentially con-
stant in the short-run, and it takes a long time for μ to change.

As a consequence, Elo reasoned that once a rating for a player becomes established,
then the only thing that can change the rating is the degree to which the player is performing
above or below his mean.[1] He proposed a simple linear adjustment that is proportional to
the player's deviation from his mean. More specifically, if a player's recent performance
(or score [2]) is S, then his old rating $r_{\text{(old)}}$ is updated to become his new rating by setting

$$r_{\text{(new)}} = r_{\text{(old)}} + K(S - \mu), \tag{5.1}$$

where K is a constant—Elo originally set $K = 10$. As more chess statistics became avail-
able it was discovered that chess performance is generally not normally distributed, so
both the USCF and FIDE replaced the original Elo assumption by one that postulates the
expected scoring difference between two players is a logistic function of the difference in
their ratings (this is described in detail on page 54). This alteration affects both μ and K
in (5.1), but the ratings are still referred to as "Elo ratings."

In 1997 Bob Runyan adapted the Elo system to rate international football (what
Americans call soccer), and at some point Jeff Sagarin, who has been providing sports
ratings for *USA Today* since 1985, began adapting Elo's system for American football.

---

[1]The assumption that performance is a normally distributed random variable and the idea of defining ratings
based on deviations from the mean is found throughout the literature. The paper by Ashburn and Colvert [6] and
its bibliography provides statistically inclined readers with information along these lines.

[2]In chess a win is given a score of 1 and a draw is given a score of $1/2$. Performance might be measured by
scores from a single match or by scores accumulated during a tournament.

In its present form, the Elo system works like this. You must start with some initial set of ratings (see page 56) for each competitor—to think beyond chess, consider competitors as being teams. Each time teams i and j play against each other, their respective prior ratings $r_{i(\text{old})}$ and $r_{j(\text{old})}$ are updated to become $r_{i(\text{new})}$ and $r_{j(\text{new})}$ by using formulas similar to (5.1). But now everything is going to be considered on a relative basis in the sense that S in (5.1) becomes

$$S_{ij} = \begin{cases} 1 & \text{if } i \text{ beats } j, \\ 0 & \text{if } i \text{ loses to } j, \\ 1/2 & \text{if } i \text{ and } j \text{ tie,} \end{cases} \qquad (5.2)$$

and μ becomes

$\mu_{ij} =$ the number of points that team i is expected to score against team j.

The new assumption is that μ_{ij} is a logistic function of the difference in ratings

$$d_{ij} = r_{i(\text{old})} - r_{j(\text{old})}$$

prior to teams i and j playing a game against each other. The *standard* logistic function is defined to be $f(x) = 1/(1 + e^{-x})$, but chess ratings employ *the base-ten version*

$$L(x) = \frac{1}{1 + 10^{-x}}.$$

The functions $f(x)$ and $L(x)$ are qualitatively the same because $10^{-x} = e^{-x(\ln 10)}$, and their graphs each have the following characteristic s-shape.
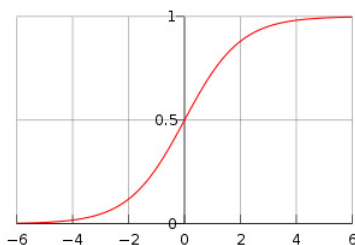

Figure 5.1 Graph of a logistic function

The precise definition of μ_{ij} for Elo chess ratings is

$$\mu_{ij} = L(d_{ij}/400) = \frac{1}{1 + 10^{-d_{ij}/400}}, \quad \text{where} \quad d_{ij} = r_{i(\text{old})} - r_{j(\text{old})}, \qquad (5.3)$$

so the respective formulas for updating ratings of teams (or players) i and j are as follows.

Elo Rating Formulas

$$r_{i(\text{new})} = r_{i(\text{old})} + K(S_{ij} - \mu_{ij}) \quad \text{and} \quad r_{j(\text{new})} = r_{j(\text{old})} + K(S_{ji} - \mu_{ji}),$$

where S_{ij} is given by (5.2) and μ_{ij} is given by (5.3).

However, before you can put these formulas to work to build your own rating system, some additional understanding of K and the 400 value in (5.3) is required.

Elegant Wisdom

The simple elegance of the Elo formula belies its wisdom in the sense that Elo inherently rewards a weaker player for defeating a stronger player to greater degree than it rewards a stronger player for beating a weak opponent. For example, if an average chess player's rating is 1500 while that of a stronger player is 1900, then

$$\mu_{avg,str} = \frac{1}{1 + 10^{-(1500-1900)/400}} = \frac{1}{11} \approx .09,$$

but

$$\mu_{str,avg} = \frac{1}{1 + 10^{-(1900-1500)/400}} = \frac{1}{1.1} \approx .91.$$

Therefore, the reward to the average player for beating the stronger player is

$$r_{avg(\text{new})} - r_{avg(\text{old})} = K(S_{avg,str} - \mu_{avg,str}) = K(1 - .09) = .91K,$$

whereas if the stronger player defeats the average player, then the reward is only

$$r_{str(\text{new})} - r_{str(\text{old})} = K(S_{str,avg} - \mu_{str,avg}) = K(1 - .91) = .09K.$$

The K-Factor

The "K-factor," as it is know in chess circles, is still a topic of discussion, and different values are used by different chess groups. Its purpose is to properly balance the deviation between actual and expected scores against prior ratings.

If K is too large, then too much weight is given to the difference between actual and expected scores, and this results in too much volatility in the ratings—e.g., a big K means that playing only a little above expectations can generate a big change in the ratings. On the other hand, if K is too small, then Elo's formulas lose their ability to account for improving or deteriorating play, and the ratings become too stagnant—e.g., a small K means that even a significant improvement in one's play cannot generate much of a change in the ratings.

Chess: The K-factor for chess is often allowed to change with the level of competition. For example, FIDE sets

$K = 25$ for new players until 30 recognized games have been completed;
$K = 15$ for players with > 30 games whose rating has never exceeded 2400;
$K = 10$ for players having reached at least 2400 at some point in the past.

At the time of this writing FIDE is considering changing these respective K-factors to 30, 30, and 20.

Soccer: Some raters of soccer (international football) allow the value of K to increase with the importance of the game. For example, it is not uncommon to see Internet sites

report that the following K values are being used.

$K = 60$ for World Cup finals;
$K = 50$ for continental championships and intercontinental tournaments;
$K = 40$ for World Cup qualifiers and major tournaments;
$K = 30$ for all other tournaments;
$K = 20$ for friendly matches.

The K-factor is one of those things that makes building rating systems enjoyable because it allows each rater the freedom to customize his/her system to conform to the specific competition being rated and to add their own personal touch.

The Logistic Parameter ξ

The parameter $\xi = 400$ in the logistic function (5.3) comes from the chess world, and it affects the spread of the ratings. To see what this means in terms of comparing two chess players with respective ratings r_i and r_j, do a bit of algebra to observe that

$$\mu_{ij} = \frac{10^{r_i/400}}{10^{r_i/400} + 10^{r_j/400}} \implies \frac{\mu_{ij}}{\mu_{ji}} = \frac{10^{r_i/400}}{10^{r_j/400}} \implies \mu_{ij} = \mu_{ji}\left[10^{(r_i - r_j)/400}\right].$$

Since μ_{ij} is the number points that player i is expected to score against player j, this means that if $r_i - r_j = 400$, then player i is expected to be ten times better than player j. In general, for every 400 rating points of advantage that player i has over player j, the probability of player i beating player j is expected to be ten times greater than the probability of player j beating player i.

Moving beyond chess, the same analysis as that used above shows that replacing 400 in (5.3) by any value $\xi > 0$ yields

$$\mu_{ij} = L(d_{ij}/\xi) = \frac{1}{1 + 10^{-d_{ij}/\xi}} = \frac{10^{r_i/\xi}}{10^{r_i/\xi} + 10^{r_j/\xi}} \tag{5.4}$$

$$\implies \frac{\mu_{ij}}{\mu_{ji}} = \frac{10^{r_i/\xi}}{10^{r_j/\xi}} \implies \mu_{ij} = \mu_{ji}\left[10^{(r_i - r_j)/\xi}\right].$$

Now, for every ξ rating points of advantage that team i has over team j, the probability of team i beating team j is expected to be ten times greater than the probability of team j beating team i. Fiddling with ξ is yet another way to tune the system to make it optimal for your particular needs.

Constant Sums

When scores S_{ij} depend only on wins, loses, and ties as described in (5.2) for chess, then $S_{ij} + S_{ji} = 1$, and (as shown below) this will force the Elo ratings to always sum to a constant value, regardless of how many times ratings are updated. This constant-sum feature also holds for competitions in which S_{ij} depends on numbers of points scored provided $S_{ij} + S_{ji} = 1$. Below is the formal statement and proof.

> ### The Elo Sums Remain Constant
>
> As long as $S_{ij} + S_{ji} = 1$, then regardless of how the scores S_{ij} are defined, the sum of all Elo ratings $r_i(t)$ at any time $t > 0$ is always the same as the sum of the initial ratings $r_i(0)$. In other words, if there are m teams, then
>
> $$\sum_{k=1}^{m} r_k(0) = \sigma \implies \sum_{k=1}^{m} r_k(t) = \sigma \quad \text{for all } t > 0.$$
>
> In particular, if no competitor deserves an initial bias, then you might assign each participant an initial rating of $r_i(0) = 0$. Doing so guarantees that subsequent Elo ratings always sum to zero, and thus the mean rating is always zero.
>
> Clearly, setting each initial rating equal to x insures that the mean rating at any point in time is always x.

Proof. Notice that regardless of the value of the logistic parameter ξ, equation (5.4) ensures that

$$\mu_{ij} + \mu_{ji} = 1.$$

This together with $S_{ij} + S_{ji} = 1$ means that after a game between teams i and j, the respective updates (from page 54) to the old ratings $r_{i(\text{old})}$ and $r_{j(\text{old})}$ are $K(S_{ij} - \mu_{ij})$ and $K(S_{ji} - \mu_{ji})$, so

$$K(S_{ij} - \mu_{ij}) + K(S_{ji} - \mu_{ji}) = 0. \tag{5.5}$$

Therefore,

$$\sum_{k=1}^{m} r_{k(\text{new})} = \sum_{k \neq i,j}^{m} r_{k(\text{old})} + [r_{i(\text{old})} + K(S_{ij} - \mu_{ij})] + [r_{j(\text{old})} + K(S_{ji} - \mu_{ji})]$$

$$= \sum_{k=1}^{m} r_{k(\text{old})}. \quad \blacksquare$$

Caution! Equation (5.5) says that the *change* in the rating for team i is just the negative of that for team j, but this does *not* mean that $r_{i(\text{new})} + r_{j(\text{new})} = 0$.

Elo in the NFL

To see how Elo performs on something other than chess we implemented the basic Elo scheme that depends only on wins and loses to rate and rank NFL teams for the 2009–2010 season. We considered all of the 267 regular-season and playoff games, and we used the same ordering of the teams as shown in the table on page 42.

While we could have used standings from prior years or from preseason games, we started everything off on an equal footing by setting all initial ratings to zero. Consequently the ratings at any point during the season as well as the final ratings sum to zero, and thus their mean value is zero.

Together with the win-lose scoring rule (5.2) on page 54, we used a K-factor of $K = 32$ (a value used on several Internet gaming sites and in line with those of the chess community), but we increased the logistic factor from 400 to $\xi = 1000$ to provide a better spread in the final ratings. The final rankings are not sensitivie to the value of ξ, and our increase in ξ did not significantly alter the final rankings, which are shown below.

Rank	Team	Rating	Rank	Team	Rating
1.	SAINTS	173.66	17.	PANTHERS	11.474
2.	COLTS	170.33	18.	NINERS	-1.2844
3.	CHARGERS	127.58	19.	GIANTS	-5.3217
4.	VIKINGS	103.50	20.	BRONCOS	-11.126
5.	COWBOYS	89.128	21.	DOLPHINS	-26.717
6.	EAGLES	69.533	22.	BEARS	-28.142
7.	PACKERS	67.829	23.	JAGUARS	-36.214
8.	CARDINALS	53.227	24.	BILLS	-53.350
9.	JETS	50.143	25.	BROWNS	-74.664
10.	PATRIOTS	39.633	26.	RAIDERS	-83.319
11.	TEXANS	33.902	27.	SEAHAWKS	-88.845
12.	BENGALS	33.012	28.	CHIEFS	-109.28
13.	RAVENS	32.083	29.	REDSKINS	-110.21
14.	FALCONS	28.118	30.	BUCS	-130.10
15.	STEELERS	27.125	31.	LIONS	-170.81
16.	TITANS	13.222	32.	RAMS	-194.12

(5.6)

NFL 2009–2010 Elo rankings based only on wins & loses with $\xi = 1000$ and $K = 32$

Since these Elo ratings are compiled strictly on the basis of wins and loses, it would be surprising if they were not well correlated with the win percentages reported in (4.18) on page 46. Indeed, the correlation coefficient (4.20) is $R_{\mathbf{rw}} = .9921$. Proceeding with linear regression as we did for the Keener ratings on pages 47 and 48 produces

$$\text{Est Win\% for team } i = .5 + .0022268\, r_i,$$

which in turn yields

$$\text{MAD} = .017958 \quad \text{and} \quad \text{MSE} = .0006.$$

While these are good results, it is not clear that they are significant because in essence they boil down to using win-lose information to estimate win percentages.

A more interesting question is, how well does Elo *predict* who wins in the NFL?

Hindsight Accuracy

While it is not surprising that Elo is good at estimating win percentages, it is not clear what to expect if we take the final Elo ratings back in time as described on page 48 to predict in hindsight who wins each game. Looking back at the winners of all of the 267 NFL games during 2009–2010 and comparing these with what Elo predicts—i.e., the higher final rated team is predicted to win—we see that Elo picks 201 game correctly, so

$$\text{Elo Hindsight Accuracy} = 75.3\%.$$

While it seems reasonable to add a home-field-advantage factor to each of the ratings for the home team in order to predict winners, incorporating a home-field-advantage factor does not improve the hindsight accuracy for the case at hand.

As mentioned in the Keener discussion on page 48, the "hindsight is 20-20" cliché is not valid for rating systems—even good ones are not going to predict all games correctly in hindsight. The Elo hindsight accuracy of 75.3% is very good—see the table of comparisons on page 123.

Foresight Accuracy

A more severe test of a ranking system is how well the vector $\mathbf{r}(j)$ of ratings that is computed after game j can predict the winner in game $j + 1$. Of course, we should expect this *foresight* accuracy to be less than the hindsight accuracy because hindsight gets to use complete knowledge of all games played while foresight is only allowed to use information prior to the game currently being analyzed.

A home-field-advantage factor matters much more in Elo foresight accuracy than in hindsight accuracy because if all initial ratings are equal (as they are in our example), then in the beginning, before there is enough competition to change each team's ratings, home-field consideration is the only factor that Elo can use to draw a distinction between two teams. When a home-field advantage factor of $H = 15$ is added to the ratings of the home team when making foresight predictions from the game-by-game ratings computed with $\xi = 1000$ and $K = 32$, Elo correctly predicts the winner in 166 out of 267 games, so

$$\text{Elo Foresight Accuracy} = 62.2\%.$$

Incorporating Game Scores

Instead of building ratings based just on wins and loses, let's jazz Elo up a bit by using the number of points P_{ij} that team i scores against team j in each game. But rather than using raw scores directly, apply the logic that led to (4.2) on page 31 to redefine the "score" that team i makes against team j to be

$$S_{ij} = \frac{P_{ij} + 1}{P_{ij} + P_{ji} + 2}. \tag{5.7}$$

In addition to having $0 < S_{ij} < 1$, this definition ensures that $S_{ij} + S_{ji} = 1$ so that S_{ij} can be interpreted as the probability that team i beats team j (assuming that ties are excluded). Moreover, if all initial ratings are zero (as they are in the previous example), then the discussion on page 56 means that as the Elo ratings change throughout the season, they must always sum to zero and the mean rating is always zero.

When the parameters $\xi = 1000$ and $K = 32$ (the same as those from the basic win-lose Elo scheme) are used with the scores in (5.7), Elo produces the following modified ratings and rankings.

Rank	Team	Rating	Rank	Team	Rating
1.	PACKERS	58.825	17.	BRONCOS	4.1006
2.	VIKINGS	55.217	18.	BENGALS	-.75014
3.	SAINTS	49.495	19.	GIANTS	-3.5097
4.	JETS	47.215	20.	DOLPHINS	-9.3122
5.	COWBOYS	43.074	21.	TITANS	-9.8351
6.	RAVENS	40.357	22.	BEARS	-16.050
7.	CHARGERS	39.974	23.	BILLS	-23.287
8.	COLTS	39.260	24.	REDSKINS	-29.039
9.	PATRIOTS	37.860	25.	CHIEFS	-34.647
10.	NINERS	33.189	26.	SEAHAWKS	-35.150
11.	TEXANS	18.447	27.	JAGUARS	-37.050
12.	FALCONS	18.387	28.	BROWNS	-47.089
13.	EAGLES	13.984	29.	BUCS	-54.373
14.	STEELERS	9.1308	30.	RAIDERS	-62.652
15.	CARDINALS	6.1216	31.	LIONS	-72.800
16.	PANTHERS	5.2596	32.	RAMS	-84.352

$$(5.8)$$

NFL 2009–2010 Elo rankings using scores with $\xi = 1000$ and $K = 32$

Hindsight and Foresight with $\xi = 1000$, K $= 32$, H $= 15$

The ratings in (5.8) produce both hindsight and foresight predictions in a manner similar to the earlier description. But unlike the hindsight predictions produced by considering only wins and losses, the home-field-advantage factor H matters when scores are being considered. If $H = 15$ (empirically determined) is used for both hindsight and foresight predictions, then the ratings in (5.8) correctly predict

194 games (or 72.7%) in hindsight,   and   175 games (or 65.5%) in foresight.

Using Variable K-Factors with NFL Scores

The ratings in (5.8) suffer from the defect that they are produced under the assumption that all games are equally important—i.e., the K-factor was the same for all games. However, it can be argued that NFL playoff games should be more important than regular-season games in determining final rankings. Furthermore, there is trend in the NFL for stronger teams to either rest or outright refuse to play their key players to protect them from injuries during the final one or two games of the regular season—this certainly occurred in the final two weeks of the 2009–2010 season.

Consequently, the better teams play below their true scoring ability during this time while their opponents appear to Elo to be stronger than they really are. This suggests that variable K-factors should be used to account for these situations.

Let's try to fix this by setting (rather arbitrarily)

$$K = 32 \quad \text{for the first 15 weeks;}$$
$$K = 16 \quad \text{for the final 2 weeks;}$$
$$K = 64 \quad \text{for all playoff games.}$$

Doing so (but keeping $\xi = 1000$) changes the Elo ratings and rankings to become those shown below in (5.9).

Rank	Team	Rating	Rank	Team	Rating
1.	SAINTS	67.672	17.	CARDINALS	1.4959
2.	VIKINGS	63.080	18.	BENGALS	1.4707
3.	COLTS	57.297	19.	PANTHERS	-3.2548
4.	PACKERS	48.227	20.	DOLPHINS	-7.6586
5.	JETS	38.781	21.	TITANS	-7.7187
6.	CHARGERS	35.864	22.	BEARS	-18.565
7.	RAVENS	35.264	23.	REDSKINS	-22.432
8.	PATRIOTS	28.496	24.	BILLS	-22.709
9.	NINERS	26.047	25.	SEAHAWKS	-29.918
10.	COWBOYS	22.742	26.	JAGUARS	-31.326
11.	TEXANS	16.289	27.	CHIEFS	-35.945
12.	EAGLES	14.492	28.	BROWNS	-51.611
13.	FALCONS	10.531	29.	BUCS	-54.044
14.	STEELERS	7.5351	30.	RAIDERS	-58.546
15.	BRONCOS	7.0388	31.	LIONS	-68.265
16.	GIANTS	6.9994	32.	RAMS	-77.329

(5.9)

NFL 2009–2010 Elo rankings using scores with $\xi = 1000$ and $K = 32, 16, 64$

As far as final rankings are concerned, these in (5.9) look better than those in (5.8) because (5.9) more accurately reflects what actually happened when all the dust had settled. The SAINTS ended up being ranked #1 in (5.9), and, as you can see from the playoff results in Figure 5.2, the SAINTS won the Super Bowl. Furthermore, the VIKINGS gave the SAINTS a greater challenge than the COLTS did during the playoffs, and this showed up in the final ratings and rankings in (5.9).
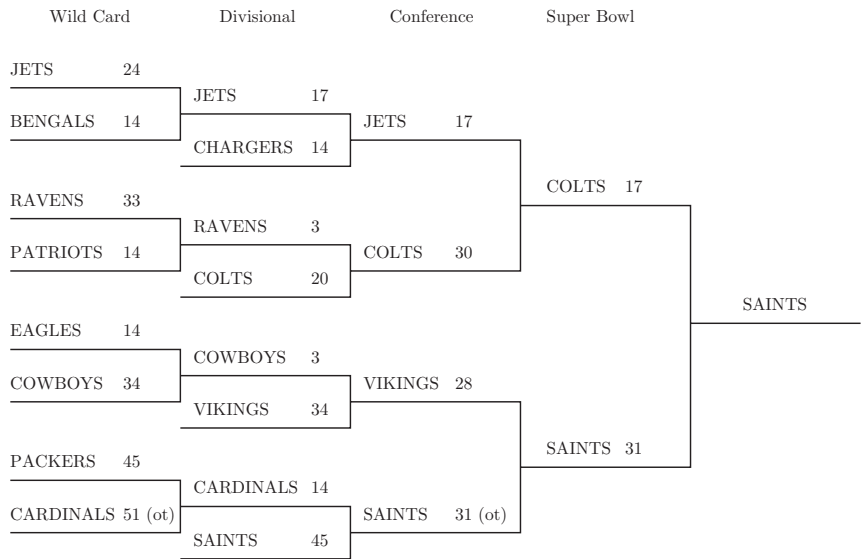
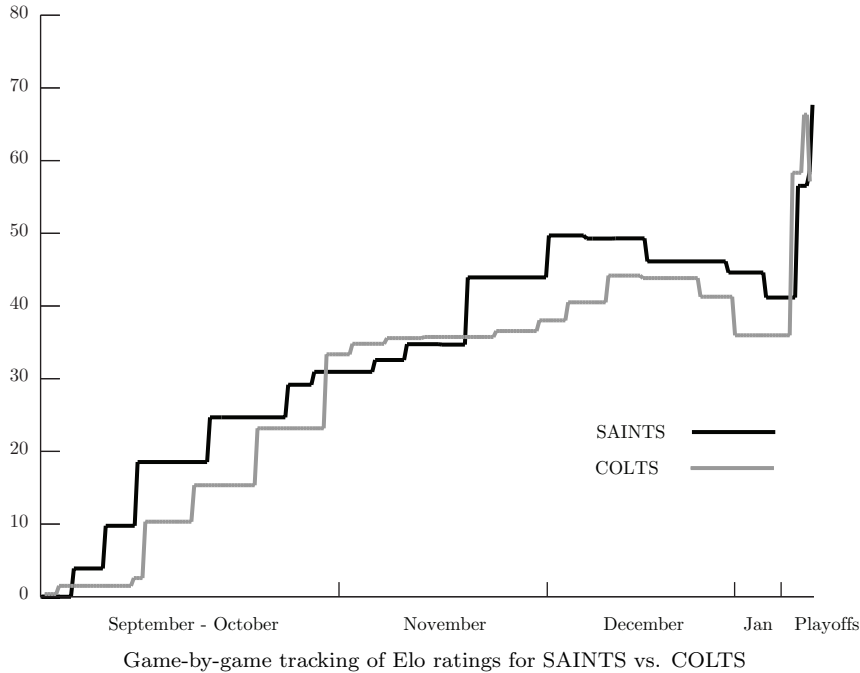| Wild Card	Divisional	Conference	Super Bowl
JETS 24			
BENGALS 14	JETS 17	JETS 17	
	CHARGERS 14		
RAVENS 33			COLTS 17
PATRIOTS 14	RAVENS 3	COLTS 30	
	COLTS 20		
EAGLES 14			SAINTS
COWBOYS 34	COWBOYS 3	VIKINGS 28	
	VIKINGS 34		
PACKERS 45			SAINTS 31
CARDINALS 51 (ot)	CARDINALS 14	SAINTS 31 (ot)	
	SAINTS 45		

Figure 5.2 NFL 2009–2010 playoff results

Hindsight and Foresight Using Scores and Variable K-Factors

The ratings in (5.9) better reflect the results of the playoffs than those in (5.8), but the ability of (5.9) to predict winners is only slightly better than that of (5.8). Changing the K-factor from a constant $K = 32$ to a variable $K = 32, 16, 64$ changes the value of the ratings, so the home-field-advantage factor H must also be changed—what works best is to use $H = 0$ for hindsight predictions and $H = 9.5$ for foresight predictions. With these values the variable-K Elo system correctly predicts
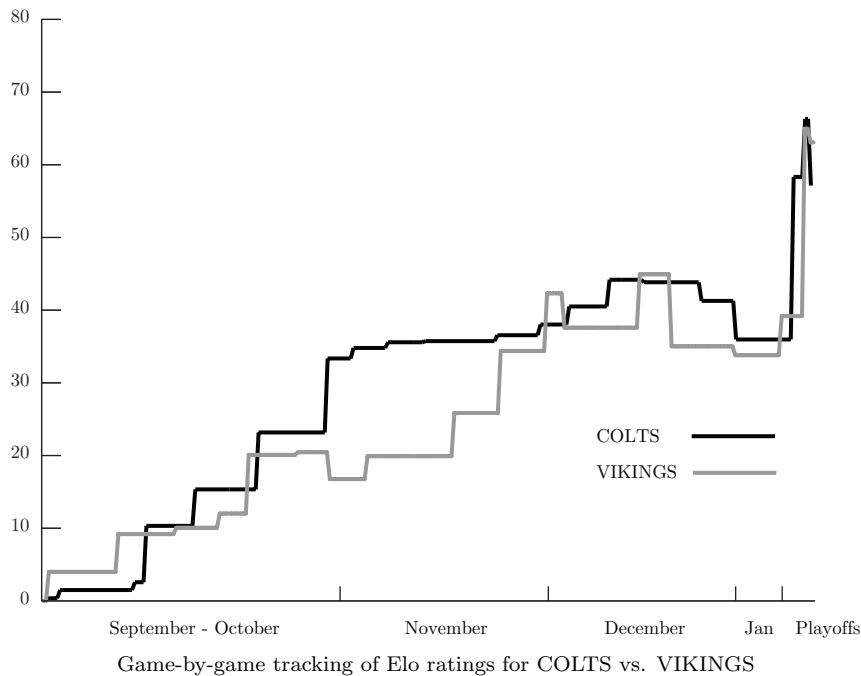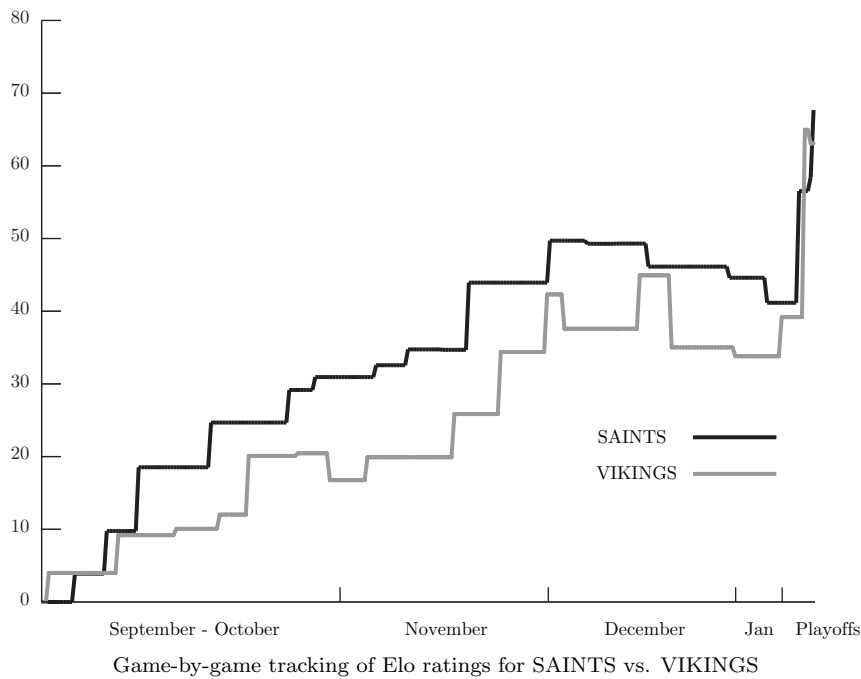
194 games (or 72.7%) in hindsight,   and   176 games (or 65.9%) in foresight.

Game-by-Game Analysis

Tracking individual team ratings on a game-by-game basis throughout the season provides a sense of overall strength that might not be apparent in the final ratings. For example, when an Elo system that gives more weight to playoff games and discounts the "end-of-season throw-away games" is used (as was done in (5.9)), it is possible that a team may have a higher final ranking than it deserves just because it has a good playoff series. For example, consider the game-by-game ratings for the three highest ranked teams in (5.9). The following three graphs track the game-by-game ratings of the SAINTS vs. COLTS, the SAINTS vs. VIKINGS, and the COLTS vs. VIKINGS.

Game-by-game tracking of Elo ratings for SAINTS vs. COLTS

Game-by-game tracking of Elo ratings for SAINTS vs. VIKINGS

Game-by-game tracking of Elo ratings for COLTS vs. VIKINGS

Notice that the SAINTS are almost always above both the VIKINGS and the COLTS, so this clearly suggests that the SAINTS were the overall stronger team throughout the 2009–2010 season. But how do the COLTS and VIKINGS compare? If the ratings for each are integrated over the entire season (i.e., if the area under the graph of each team in

the last graph above is computed), then the COLTS come out ahead of the VIKINGS. In other words, in spite of the fact that VIKINGS are ranked higher than the COLTS in the final ratings, this game-by-game analysis suggests that the COLTS were overall stronger than the VIKINGS across the whole season. As an unbiased fan who saw just about every game that both the COLTS and VIKINGS played during the 2009–2010 season (yes, I [C. M.] have the *Sunday Ticket* and a DVR), I tend to agree with this conclusion.

Conclusion

The Elo rating system is the epitome of simple elegance. In addition, it is adaptable to a wide range of applications. Its underlying ideas and principles might serve you well if you are trying to build a rating system from the ground up.

Regardless of whether or not there was an actual connection between Elo and Facebook, the following aside helps to underscore the adaptability of Elo's rating system.

ASIDE: Elo, Hot Girls, and *The Social Network*

Elo made it into the movies. Well, at least his ranking method did. In a scene from the movie *The Social Network*, Eduardo Saverin (co-founder of Facebook) is entering a dorm room at Harvard University as Mark Zuckerberg (another co-founder of Facebook) is hitting a final keystroke on his laptop, and the dialog goes like this.

Mark: "Done! Perfect timing. Eduardo's here, and he's going to have the key ingredient."

(After comments about Mark's girlfriend troubles)

Eduardo: "Are you alright?"

Mark: "I need you."

Eduardo: "I'm here for you."

Mark: "No, I need the algorithm used to rank chess players."

Eduardo: "Are you okay?"

Mark: "We're ranking girls."

Eduardo: "You mean other students?"

Mark: "Yeah."

Eduardo: "Do you think this is such a good idea?"

Mark: "I need the algorithm—I need the algorithm!"

Eduardo proceeds to write the Elo formula (5.3) from page 54 on their Harvard dorm room window. Well, almost. Eduardo writes the formula as

$$E_a = \frac{1}{1 + 10(R_b - R_a)/400}$$

instead of

$$E_a = \frac{1}{1 + 10^{(R_b - R_a)/400}} = \frac{1}{1 + 10^{-(R_a - R_b)/400}}.$$

We should probably cut the film makers some slack here because, after all, Eduardo is writing

on a window with a crayon. The film presents Elo's method as something amazingly complicated that only genius nerds appreciate, which, as you now realize, is not true. Moreover, the film explicitly implies that Elo's method formed the basis for rating Harvard's hot girls on Zuckerberg's Web site Facemash, which was a precursor to Facebook. It makes sense that Zuckerberg might have used Elo's method because Elo is a near-perfect way to rate and rank things by simple "this-or-that" pairwise comparisons. However, probably no one except Zuckerberg himself knows for sure if he actually employed Elo to rate and rank the girls of Harvard. If you know, tell us.

By The Numbers —

2895 = highest chess rating ever.
— Bobby Fischer (October 1971).

2886 = second highest rating ever.
— Gary Kasparov (March 1993).

— `chessbase.com`