An innovative approach to National Football League standings using optimal bonus points

This Draft: March 2009

Niven Winchester
Department of Economics
University of Otago
Dunedin
New Zealand

niven.winchester@otago.ac.nz

Raymond T. Stefani
College of Engineering
California State University, Long Beach
Long Beach, CA 90840
USA

stefani@csulb.edu

Abstract

Bonus points provide a simple way to improve the accuracy of league standings. We investigate the inclusion of bonuses in the National Football League (NFL) using a prediction model built on league points. Both touchdown-based and narrow-loss bonuses are shown to be significant. Our preferred system awards four points for a win, two for a tie, one point for scoring four or more touchdowns and one point for losing by seven or fewer points. Such a system would also make it easier for supporters to identify playoff contenders and place importance on otherwise meaningless end-of-game plays.

Key words: tournament design, sports predictions, NFL

JEL codes: C61, L83

1. Introduction

Sports ranking systems (SRSs) assign values to teams or individuals so that competitors may be placed in an ordinal rank from best to worst.[1] Many sport federations take great care to construct SRSs that are sensitive to the outcome of each completion so as to place each outcome into proper perspective. For example, in international soccer, the International Federation for Association Football awards more points for defeating a higher-ranked team than a lower-ranked team. The International Aeronautical Federation sponsors an SRS for hang gliding and paragliding in which each competitor's performance is evaluated as the product of factors which measure the finishing time relative to the other pilots, the quality and number of the participants and the number of skills tested.

In contrast, National Football League (NFL) teams are ranked according to win-tie-loss records, where a tie is worth half a win. Team ratings based on win-tie-loss records are less accurate that ratings that utilize additional characteristics. For example, Berry (2003) analyzed 15,728 American college football games played over a 32-year period. An optimal system based on game outcome correctly predicted the winner in 63.1% of games. An optimal system based on score difference was correct in 71.8% of matches.

Bonus points provide a way to improve the accuracy of league standings that is easy for stakeholders to understand. There are at least two examples of bonus points in top-level club league competition. In the National Hockey League (NHL), two points are awarded for a win and one bonus point is awarded for an overtime loss. A team with an overtime loss is thus more favorably considered than a team losing in regulation.

[1] See Stefani (1997 & 1999) for reviews of SRSs.

In most domestic and international rugby union competitions, four points are allocated for a win, two points for a tie and there are two types of bonuses. One bonus point is given for a team scoring four or more tries (touchdowns) and one bonus point is given for a team losing by seven or fewer points, the rugby union analogy to the NHL overtime-loss bonus.

Anecdotal evidence suggests that a rugby union-style bonus point system would assist the identification of strong teams in the NFL. Specifically, Cincinnati and Pittsburgh both finished the 2005 regular season with identical win-tie-loss records, but Cincinnati was ranked ahead of Pittsburg and awarded an easier playoff schedule based on a superior within-division record. However, during the regular season, Cincinnati lost one match by seven or fewer points while Pittsburgh recorded four losses by seven or fewer points, suggesting that Pittsburgh should be ranked above Cincinnati. Indeed, bookmakers installed Pittsburgh as three-point favorites when they visited Cincinnati in the first round of the playoffs, a game won by Pittsburgh on their way to claiming Super Bowl XL.

Until recently, lacking in the literature was an objective evaluation of current ratings methods and a discussion of ad-hoc and optimal bonus point allocations which add greater specificity to game outcome.[2] Winchester (2008) determined optimal bonus points for rugby union by regressing the home team's net score on home advantage and a combination of previous-season and current-season league points earned by the two opponents. He found that (modified) try and narrow loss bonuses were statistically significant.

We advance Winchester's (2008) methodology by specifying several alternative strength measures and quality-adjusting league points to account for the

[2] See Szymanski (2003) for a review tournament design issues.

unbalanced nature of NFL schedules. To our knowledge, we are also the first to consider bonus points in the NFL. Our analysis determines both optimal partitions and values for bonuses. We find that bonuses should be awarded for scoring four or more touchdowns and losing by seven or fewer points. We also conclude that bonus points will increase spectator interest and make it easier for supporters to identify teams in playoff contention.

The paper has five further sections. The next section provides an overview of the NFL. Section III outlines our modeling framework. Results are presented in Section IV. NFL standings for recent seasons when optimal bonus points are included are discussed in Section V. The final section concludes.

II. The National Football League

Fourteen teams formed the precursor to the NFL in 1920, the American Professional Football Association. The rules adopted by that league evolved from US college football rules. Since 1920, the currently-named NFL merged with other leagues and added other teams to expand into the current 32-team structure with two 16-team conferences, each with four, four-team divisions. Each of the 32 teams plays a regular season schedule of 16 games. In each season, a given team plays each of the other three teams in the same division twice, once at home and once away, accounting for six games. Six games are played against one-half of the other 12 teams in the same conference. Another four games are played against one-fourth of the 16 teams in the other conference. Clearly, the schedule is unbalanced except against teams in the same division.

If a regulation game ends in a tie, one overtime period is played where the first team to score wins the game. Few ties remain – there was only one tied game in ten seasons from 1998 through 2007, with 256 games currently played during each regular season.

At the conclusion of the regular season, 12 teams move forward into playoff competition, six from each conference, consisting of the four divisional champions from each conference and two “wild-card” runner-up teams from each conference. The deciding factor is league standings based on wins (and infrequent ties) with a complex tie-breaker procedure when needed. The 12 teams proceed to a ladder-type playoff terminating with the “Super Bowl” played by the two conference champions. Given the huge amounts of money at stake, it is important to evaluate the significance of current league rankings and to provide bonus points to improve significance.

There is a huge financial incentive to maintain the competitive viability of the NFL. The average NFL player earned $1.4 million in 2005 (Weisman, 2006). These salaries are made possible by TV rights of about $4 billion per year (El-Bashir and Heath, 2006). The gate receipts for the Super Bowl were estimated at $40 million (Associated Press, 2005). A 30-second Super Bowl advertisement was estimated to cost $2.6 million (La Monica, 2007). If the NFL were to adopt a bonus point system such as that commonly used in rugby union, the last minutes of an otherwise decided game could become more important due to one or both teams possibly gaining bonus points. Ensuing fan interest would increase TV viewers at game’s end. Further, bonuses will make it easier for stakeholders to identify teams with a chance of making the playoffs. Specifically, due to the highly discrete nature of current NFL ratings, a number of tie-breaking procedures are commonly required to separate teams prior to the playoffs. These procedures may involve comparing within-division win-tie-loss records, win-tie-loss records in common games, or opponents’ win-tie-loss records (a “strength of schedule” measure). Thus, supporters may be required to process a large amount of information and, in the case of strength of schedule calculations, consider a large number of alternatives to determine playoff contenders. Although a bonus points system may also require tie breakers, they would be needed less frequently than under the current system.

III. Modeling Framework

We determine optimal bonus points using a prediction model. NFL prediction models in the public domain are presented by Goode (1976), Leake (1976), Stefani (1977, 1980, 1987 & 1998), Harville (1980), Zuber *et al.* (1985), Glickman and Stern (1998), and Boulier and Stekler (2003). These studies create ratings by observing factors such as match scores and yards per pass. Accurate NFL predictions result from the predictive systems in the same way that higher seeded competitors win a large percentage of matches in seeded tournaments. What is lacking is an analysis of the ability of league standings to correlate with the success of higher-ranked teams in subsequent games and the consideration of a bonus point system that could improve the selectivity of NFL standings. Our analysis addresses these shortcomings.

We determine optimal bonus points by including touchdown and narrow-loss bonuses as endogenous explanatory variables in a prediction model following Winchester (2008). We use the prediction model to choose (a) the value of a touchdown bonus and the minimum number of touchdowns required to earn this bonus, and (b) the value of a narrow-loss bonus and the minimum losing margin required to earn this bonus. Everything else held constant, a strong team should beat a weak team, so choosing bonuses to maximize prediction accuracy selects the allocation of bonuses that provides the best measure of team strength.

We do not account for changes in player and/or coach behavior induced by bonus points. Notably, a touchdown bonus may induce coaches to choose a passing or running play instead of a kicking play. However, as winning attracts the greatest reward in our bonus point specification, we speculate that introducing bonuses will not stimulate major tactical changes. Investigating this conjecture presents an interesting avenue for further research, perhaps building on Romer (2006).

As is common in prediction models, we use the home team's winning margin (points scored by the home team minus points scored by the away team) to characterize the outcome of a match. Important determinants of match outcomes include home advantage and the strength of the two opponents. We prefer a small number of home advantage parameters. Suppose that one home advantage is averaged over all league games played in a double round-robin regular season competition among N teams. Since $N(N-1)$ games are played, the variance of that average home advantage is inversely proportional to $N(N-1)$. If a separate home advantage is found for each of the N teams by using home-away paired results, $N-1$ sets of paired games are used per parameter, increasing variance by a factor of N. As fewer home advantage parameters are found, more games are used per parameter and variance is reduced. Additionally, Harville and Smith (1994), Clarke (1993 & 2005) and Winchester (2008) find that it is not necessary to include a large number of parameters to model home advantage.[3]

We specify two home advantage parameters. We allow home advantage for intra-division contests to differ from that for inter-division matches. We rationalize this on the grounds that teams play opponents within their division more frequently

[3] By specifying a small number of home advantage parameters, we are unable to include fixed effects to capture unobservable characteristics. However, we included team fixed effects and year dummies in unreported estimations. Our findings are unaltered by the inclusion of these parameters.

than opponents outside their division and, in general, a division is made up of teams in relatively close geographic proximity. These factors result in less ground familiarity and greater travel fatigue for the away team in inter-division than intra-division matches. Home advantage may also depend on other factors, such as whether a match is played on grass and the away team usually plays on turf. We do not consider such factors.

We estimate a team's strength by calculating wins per game, the number of times a large number of touchdowns were scored per game, and narrow losses per game. League points are awarded for winning, tieing, scoring a large number of touchdowns and losing narrowly. We consider three specifications for calculating average league points.

- TV: A time-varying weighted average of league points in the previous and current seasons.
- MA: A moving average of league points.
- EXP: An exponentially-weighted average of league points.

Our strength calculations extend Winchester's (2008) methodology in three ways. First, we consider three approaches for measuring team strength, whereas Winchester only models a TV specification. Second, Winchester employs an exogenous across-season weight, while we allow the data to determine this weight in our TV specification. Third, since there can be large differences in schedule difficulty across teams, we compute quality-adjusted league points (QALPs) to improve accuracy. We do this by calculating ratings for each team in each time period using an exponential smoothing model following Clarke (1993). [4]

[4] Note that we use exponential calculations in two settings. First, we use exponential smoothing to generate team ratings to quality adjust league points. Second, we calculate exponential-weighted league points as one of three methods to determine optimal bonus points.

A. Team Ratings using Exponential Smoothing

We generate team ratings to quality adjust league points using a prediction model based on winning margins. Ratings are updated in an exponential fashion according to whether a team performs better or worse than expected. Specifically,

$$PM_{ij,t} = \beta^{H} + \beta^{D} D_{ij,t} + R_{i,t-1} - R_{j,t-1} \tag{1}$$

where $PM_{ij,t}$ is the predicted winning margin of home team i against road team j in cumulative week t (cumulative weeks are not reset at the beginning of each year), β^H is home advantage applying to all contests; β^D is additional home advantage for inter-division contests; D_{ij} is a binary variable equal to one if i and j compete in different divisions, and R_i is the rating of team i.

Team i's rating is increased (decreased) if i performs better (worse) than expected according to the following equation:

$$R_{i,t} = \delta E_{ij,t-1} + R_{i,t-1} \tag{2}$$

where δ is the smoothing constant and $E_{ij,t}$ is the "prediction error".

Following Clarke (1993, p. 756) the error function in our exponential smoothing algorithm uses a power function "to reduce the relative errors of matches with large actual or predicted margins, and to increase the weighting across the 'win-loss' boundary." Accordingly,

$$E_{ij,t} = sign(M_{ij,t}).\left|M_{ij,t}\right|^{\rho} - sign(PM_{ij,t}).\left|PM_{ij,t}\right|^{\rho} \tag{3}$$

where M_{ij} is home team i's winning margin against road team j and ρ is a parameter to be estimated.

The model is a nonlinear program with discontinuous derivatives. We code the model using the General Algebraic Modeling System (GAMS) and determine values for β^H β^D, δ and ρ to minimize the sum of squared prediction errors using the solver MINOS. We estimate the model for the 2002-2007 seasons.

The model requires the assignment of initial ratings. We generate initial ratings for 2003 by fitting the model to 2002 data. Starting values are generated by choosing ratings at the beginning of the 2002 season based on points scored for and against in the preceding season and estimating the model using 2002 data. Since Houston first played in the NFL in 2002, we set the initial rating for this team equal to the lowest initial rating observed for the other 31 teams.

Ratings at the beginning of the 2002 season are scaled so that the average rating across teams equals 100. The model is then re-estimated for 2003-2007 using ratings at the end of the 2002 season as initial ratings. This procedure reduces the impact of exogenously-chosen values on estimated parameters.

B. Predictions using Quality-Adjusted League Points

The quality-adjusted win for team i in week t, $W^*_{i,t}$, is given by:

$$W^*_{i,t} = W_{i,t} R^o_{i,t} \tag{4}$$

where $W_{i,t}$ is equal to one if, in week t, i won, 0.5 if i tied and zero if i lost, and R_i^o is the rating of i's opponent estimated from (1). Similarly, we calculate quality-adjust touchdown bonuses ($TD^*_{i,t}$) and narrow-loss bonuses ($L^*_{i,t}$) as:

$$TD^*_{i,t} = T_{i,t} R^o_{i,t} \tag{5}$$

$$L^*_{i,t} = L_{i,t} R^o_{i,t} \tag{6}$$

where $TD_{i,t}$ and $L_{i,t}$ equal one if team i, respectively, earned a touchdown bonus or a narrow-loss bonus (and zero otherwise).

Combining (4), (5) and (6) yields an expression for the number of QALPs earned by team i in week t, $P^*_{i,t}$:

$$P^*_{i,t} = \theta^{WIN} W^*_{i,t} + \theta^{TD} TD^*_{i,t} + \theta^{LOSS} L^*_{i,t} \tag{7}$$

where θ^{WIN}, θ^{TD} and θ^{LOSS} are competition points awarded for, respectively, winning, scoring more than a certain number of touchdowns, and losing by less than a certain number of points.

Our prediction model used to determine optimal bonuses is based on the following equation:

$$M_{ij,t} = \beta^H + \beta^D D_{ij,t} + \beta^S (S^q_{i,t-1} - S^q_{j,t-1,y}) + \varepsilon_{ij,t} \tag{8}$$

where S_i^q is the strength of team i calculated using method q (q = TV, MA, EXP), β^S captures the influence of strength differences on the home team's winning margin, and ε is an error term.

B.1. TV Specification

Strength in our TV specification is a time-varying weighted average of competition points from the previous and current seasons. This necessitates an alternative time index to that used above. We let k identify weeks (which are reset at the beginning of each year) and y identify years. The strength of team i in week k of year y is:

$$S_{i,k,y}^{TV} = (1-\lambda_{i,k,y})\overline{P}_{i,k,y} + \lambda_{i,k,y}\overline{P}_{i,17,y-1} \tag{9}$$

where $\overline{P}_{i,k,y}$ indicates competition points earned per-game by team i in year y at the completion of week k, and λ_{iy} is the weight on competition points earned in the previous season. We replace k with 17 when referring to competition points earned in the previous year as there are 17 weeks in each season.

In the first game of each season, the weight on competition points from the previous season is one. This weight decreases by a constant fraction each game until the number of games played in the current season equals or exceeds the number of games included in strength calculations. Once this happens, strength measures only include outcomes from the current season. Formally, let N^{TV} denote the number of games included in strength calculations. If $g_{i,k,y}$ is the number of games played by team i in year y at the completion of week k, the weight on competition points earned in the previous season ($\lambda_{i,k,y}$) is:

$$\lambda_{i,k,y} = \begin{cases} (N^{TV} - g_{i,k,y})/N^{TV} \text{ if } \mathrm{g_{i,k,y}} \leq N^{TV} \\ 0 \text{ if } g_{i,k,y} > N^{TV} \end{cases} \tag{10}$$

If $g_{i,k,y} < N^{TV}$ per game wins, touchdowns and narrow-loss bonuses are calculated using all games played by team i in year y, otherwise i's most recent N^{TV} games are used to calculate averages. Substitution of (9) into (8) and appropriately modifying the time index yields the equation to be estimated. Expanding this equation using (7) yields:

$$\begin{aligned} M_{ij,k,y} &= \beta^{H} + \beta^{D} D_{ij,k,y} \\ &+ \beta^{S}.\theta^{WIN}(\lambda_{i,k,y}\overline{W}_{i,17,y-1} + (1-\lambda_{i,k,y})\overline{W}_{i,k,y} - \lambda_{j,k,y}\overline{W}_{j,17,y-1} - (1-\lambda_{j,k,y})\overline{W}_{j,k,y}) \\ &+ \beta^{S}.\theta^{TD}(\lambda_{i,k,y}\overline{TD}_{i,17,y-1} + (1-\lambda_{i,k,y})\overline{TD}_{i,k,y} - \lambda_{j,k,y}\overline{TD}_{j,17,y-1} - (1-\lambda_{j,k,y})\overline{T}_{j,k,y}) \\ &+ \beta^{S}.\theta^{LOSS}(\lambda_{i,k,y}\overline{L}_{i,17,y-1} + (1-\lambda_{i,k,y})\overline{L}_{i,k,y} - \lambda_{j,k,y}\overline{L}_{j,17,y-1} - (1-\lambda_{j,k,y})\overline{L}_{j,k,y}) \\ &+ \varepsilon_{ij,k,y} \end{aligned} \tag{11}$$

where $\overline{W}_{i,k,y}$, $\overline{TD}_{i,k,y}$ and $\overline{L}_{i,k,y}$ denote, respectively, the average number of quality-adjusted wins, touchdown bonuses and narrow-loss bonuses earned by team i in year y at the completion of week k.

Since an allocation of competition points is invariant to multiplication by a positive scalar, we normalize competition points with respect to θ^{WIN}. That is, we set $\theta^{WIN} = 1$ and express values for θ^{TD} and θ^{LOSS} relative to the number of competition points awarded for a win. Parameters to be estimated include β^{H}, β^{D}, β^{S}, θ^{TD} and θ^{LOSS}. The best rating system could be defined as that producing the least squared error when predicting future score difference or, alternatively, the system for which the most winning teams are predicted. The latter strategy operates mainly upon 0-1 data in which winning and losing is often decided by a few key successful plays, key turnovers and/or key refereeing decisions. Conversely, score differences offer more than an additional order of magnitude of levels distinguishing the performance of two teams. Parameters which optimize predicted score difference over many games can be expected to provide a much more robust model which is not as strongly influenced by luck as a model optimized for correctly predicted wins. The former approach is used herein to order the relative success of various models. We estimate the model using nonlinear least squares (NLS).

B.2. MA and EXP Specifications

Our MA and EXP specifications do not require us to distinguish data from different years, so we return to using t to index (cumulative) weeks. The strength of team i in week t in our moving average specification, $S_{i,t}^{MA}$, is:

$$S_{i,t}^{MA} = \theta^{WIN}\overline{W}_{i,t} + \theta^{TD}\overline{TD}_{i,t} + \theta^{LOSS}\overline{L}_{i,t} \tag{12}$$

where the average number of quality-adjusted wins, touchdown bonuses, and narrow-loss bonuses are calculated using each team's most recent N^{MA} matches.

Substituting (12) into (7) yields an expression similar to (11). As for our TV specification, we estimate β^H, β^D, β^S, θ^{TD} and θ^{LOSS} using NLS. We consider TV and MA specifications that assign a higher weight to league points from more recent matches than league points from less recent matches in unreported simulations. Results for the weighted specifications are similar to those for the specifications described below.

When an exponential weight is used, the strength of team i in week t, $S_{i,t}^{EXP}$, is given by:

$$S_{i,t}^{EXP} = \eta(\theta^{WIN} W_{i,t}^{*} + \theta^{TD} TD_{i,t}^{*} + \theta^{LOSS} L_{i,t}^{*}) + (1-\eta) S_{i,t-1}^{EXP} \qquad (13)$$

where η is a parameter to be estimated, $0 \le \eta \le 1$.

Our EXP specification is a nonlinear program. We code the model using GAMS and choose values for η, β^H, β^D, β^S, θ^{TD} and θ^{LOSS} to minimize the sum of squared prediction errors using the solver CONOPT. Initial strength needs to be assigned to estimate the EXP specification. For each team, we set initial values for W^*, TD^* and L^* equal to, respectively, the average number of wins, touchdown bonuses and narrow-loss bonuses earned in the 2002 regular season (the season prior to the commencement of our predictions).

IV. Model Estimation and Results

As discussed above, our sample includes regular season matches from the 2002-2007 seasons. We select these years as the same teams competed over that period. We source data from www.nfl.com. Since some predictions require results lagged by one year, we estimated (11) for the 2003-2007 seasons. Several matches were played at irregular venues during this period and required special treatment. We regarded New Orleans' home opener against the New York Giants played at Giants Stadium due to Hurricane Katrina as an away match for New Orleans. Furthermore, we treated New Orleans' other 2005 home games played at either the Alomodome (San Antonio, Texas) or Tiger Stadium (Baton Rouge, Louisiana) as games played at neutral venues. We also treated the two matches in our sample played outside the US

(Arizona versus San Francisco in week four, 2005, and Miami versus New York Giants in week eight, 2007) as neutral games. We set $\beta^H = \beta^D = 0$ for games played at neutral venues.

Table 1 displays game scores, touchdowns per game and win percentages for home and away teams averaged over the 2003-2007 seasons. Calculations are included separately for (a) intra-division, (b) inter-division, and (c) all matches. Statistics for all matches indicate that home teams score about two and a half more points than away teams on average, and around 58% of matches are won by the home team. Comparing panels (a) and (b) reveals that the average home winning margin is around one point more and the home win percentage five percentage points higher for inter-division matches than intra-division matches. Home teams also score more touchdowns when hosting inter-division rivals as opposed to intra-division opponents.

A. Ratings used to Quality Adjust League Points

Estimated parameter values for our exponential smoothing model built on winning margins are $\beta^H = 1.72$, $\beta^D = 1.30$, $\delta = 0.23$ and $\rho = 0.61$. The model correctly selected the winning team in 848 of 1,280 games, 66.3% accuracy.

The model produces ratings for each team in each week of each season. Average ratings across years by team, conference and division are shown in Table 2. New England (107.4) and Indianapolis (106.9) attained the highest average ratings over the five-year period. New England was undefeated in the 2007 regular season and achieved an average rating of 113.1, which is the highest year-average rating. The lowest rated teams according to our calculations are Arizona and San Francisco, which attained average ratings of 94.6 and 94.8, respectively, over our five-year sample. The results also suggest that the AFC-East, which includes New England, is the strongest division and the NFC-South, which includes Arizona and San Francisco, the weakest. Additionally, our numbers indicate that, at a neutral field, an AFC team will beat an NFC team by 2.2 (101.1 – 98.9) points on average. We scale exponential smoothing ratings so that the average rating across teams is equal to one when calculating quality-adjusted league points.

B. TV, MA and EXP Specifications

In each specification, we vary the minimum number of touchdowns required for a touchdown bonus from one to seven and the minimum losing margin required for a narrow-loss bonus from one to 12. In our TV and MA specifications, we also vary the number of games included in strength calculations (N^{TV} and N^{MA} respectively) from one to 16. We choose bonus partitions and values for N^{TV} and N^{MA} to minimize the sum of squared prediction errors.

Results for each specification when bonuses are (a) excluded and (b) included are presented in Table 3. The model explains 16%-17% of the variation in the sum of squared net scores and correctly predicts the winning team in around 64% of the games. The results for home advantage parameters are consistent across specifications and suggest that within-division home advantage is around two points and has a *p*-value less than 0.01. Additional home advantage for inter-division matches is around 1.4 points, and is significant at a 10% significance level but not a 5% significance level.

Twelve games are included in strength calculations in our TV specification regardless of the treatment of bonuses. As a result, the weight on current season results increases by 1/12 after each game and only results from the current season are included in strength calculations from the thirteenth game onwards. In our MA specification, competition points are averaged over 14 games when bonuses are excluded and 13 games when bonuses are included. The relative weight on more recent results is determined by η in our EXP specification, which is equal to 0.11 (and is not reported in Table 3).

Estimates for β^S indicate that, other factors held constant, a team that wins every game will beat a team that loses every game by around 20 points in all specifications. The optimal touchdown partition is three in our EXP specification and four in other specifications. The optimal narrow loss partition is eight in all models. Significantly, *p*-values for bonus parameters, with one exception, are less than 0.01 in all models. The *p*-value for the touchdown bonus is not less than 0.01 in or EXP specification, but it is below 0.05.

Including bonuses also increases the number of correct predictions in all specifications. The inclusion of bonuses is most beneficial in the TV specification, where including bonuses increases the R-squared by around 10 percent and the number of correct predictions by 13. Including bonuses increases the number of correct predictions by a larger number when bonus partitions are selected to maximize the number of correct predictions. For example, there are 34 additional correct predictions from the inclusion of bonus points in our MA specification when bonus partitions maximize the number of correct predictions.

Estimates for the touchdown bonus range from 0.35 to 0.53 and those for the narrow-loss bonus from 0.45 to 0.49. Like our optimal bonus partitions, these values seem reasonable, although a little high if 0.5 points are awarded for a tie. Overall, the results in Table 3 show that an allocation of competition points that includes narrow-loss and touchdown bonuses is better at revealing strong teams than an allocation that only rewards wins and ties.

Both bonuses should equal a common value that is easily converted to a small integer so that the system is easy to understand. Two possible values include ⅓ and ¼. Using ⅓ (and equating a tie to half a win) results in a win-tie-touchdown-narrow loss allocation of 6-3-2-2. Using ¼ produces a 4-2-1-1 system. The data cannot reject restricting bonuses to these values (p-values for joint tests that $\theta^{TD} = \theta^{LOSS} = ⅓$ are greater than 0.3 and p-values for the restriction $\theta^{TD} = \theta^{LOSS} = ¼$ are greater than 0.1). Although bonuses are closer to optimal values under 6-3-2-2, we favour 4-2-1-1 for its simplicity. Additionally, the status quo only rewards wins (and ties) and the relative value of a win is higher under 4-2-1-1 than 6-3-2-2. Consequently, there will be less resistance to 4-2-1-1 than 6-3-2-2 from stakeholders with a conventional view to league ratings.

We estimate each specification when $\theta^{TD} = \theta^{LOSS} = ¼$ for a range of bonus partitions. Results (not reported) vary across specifications but there is clear evidence that the optimal touchdown partition is three or four and the optimal narrow-loss partition is seven or eight. We choose a touchdown partition of four as earning a touchdown bonus is more difficult under this cut-off than a partition of three. This makes our system more marketable to individuals with an orthodox view to league standings.

A narrow-loss partition of eight implies that a bonus may be awarded if an additional maximum scoring play (a touchdown plus a two-point conversion) by the losing team would have tied the game. Conversely, a losing team awarded a narrow-loss bonus must be within an additional maximum scoring play of winning when a partition of seven is used. For this reason, in our opinion, a loss by seven or fewer points is a superior definition of a narrow loss than a loss by eight or fewer.

In summary, similar to the system currently used for rugby union, our preferred allocation of competition points awards four points for a win, two points for a tie, one point for scoring four or more touchdowns, and one point for losing by seven or fewer points.

V. Alternative NFL Standings

As mentioned earlier, the NFL playoffs are contested by 12 teams using a ladder structure: the four division champions and the two highest-ranked non-division champions (known as the wildcard qualifiers) in each conference. In each conference, division champions are seeded one through four and the two wildcard qualifiers are seeded five and six. The number of regular season wins added to half the number of ties is used to seed teams, and there are a number of tie-breaking rules to separate teams with an equal number of such wins. In the playoffs, in each conference, seeds one and two receive a bye directly to the second round while the three and four seeds are awarded home advantage in the first round. The wild-card qualifiers must be the away team for any match played prior to the championship game, termed the Super Bowl. Each match results in one team being eliminated. Therefore, regular season rankings influence the difficulty of a team's playoff schedule in addition to determining which teams make the playoffs.

We evaluate the impact of including bonus points on NFL standings for the 2003-2007 seasons by identifying four playoff classifications and evaluating whether or not including bonuses altered a team's classification. In each conference, our playoff classification identifies (i) one and two seeds (z^*), (ii) three and four seeds (z), (iii) wildcard qualifiers (y), and (iv) teams that did not qualify for the playoffs (n). Rankings when bonus points are included are determined by replacing the number of wins with competition points in ranking decision rules currently used in the NFL. Cases where the ranking methods produced different classifications are reported in Table 4. The first letter in each cell denotes the classification using a conventional system and the second reports the classification when bonus points are included. Wins, touchdown bonuses, narrow-loss bonuses and competition points (in parentheses) accumulated by each team are displayed below classification change indicators. For example, in 2003, Dallas gained wildcard qualification but would not have qualified for the playoffs if bonus points were included. Dallas also accumulated ten wins, one touchdown bonus and one narrow-loss bonus for a total of 42 competition points.

Bonus points change playoff classifications for 17 teams in the five-year period and there are seven direct playoff classification exchanges due to the inclusion of bonuses (e.g., Dallas' wildcard qualification is exchanged for Minnesota's non-qualification in 2003). Four direct exchanges are between wildcard qualifiers and non-qualifiers, two involve swaps between three or four seeds and wildcard qualifiers, and one involves a swap between a one or two seed and a three or four seed. These observations indicate that bonuses have the largest impact on teams close to the qualification-non-qualification boundary. In other words, bonuses have little impact on teams that are clearly dominant but are significant when separating teams that are tightly bunched. This is a desirable quality.

A team with fewer wins is rated higher than a team with more wins on four occasions. Although such ratings are likely to be controversial amongst NFL supporters and administrators, using multiple performance measures (and not just the number of wins) to evaluate team strength is common elsewhere. A notable example concerns bookmakers. If strength estimates are based on the number of wins and match location is a tie breaker, the team with the highest win percentage will be the favorite, or the home team will be the favorite if two opponents have the same win percentage. The bookmakers' favorite does not concur with this decision rule in 22.1% of matches during the 2003-2007 seasons (excluding the first week of each season, when win percentages are equal across teams). The corresponding figure when only weeks five to fifteen are included (so as to minimize the impact of small sample variability and end-of-season matches that may have little impact on final league standings) is 21.3%. Although a team with fewer wins may be favored due to strength of schedule differences or the absence (or return) of key personnel, it appears that bookmakers consider a team with fewer wins to be stronger than its opponent on some occasions independent of these factors.

An interesting ranking reversal concerns St Louis and Carolina in 2004. St Louis earned wildcard qualification and Carolina finished third in its division and did not qualify for the playoffs in this year. However, our preferred allocation of competition points suggests that Carolina was the second strongest team in its division and should have qualified for the playoffs at the expense of St Louis. Although Carolina finished with a 7-9 win-loss record and St Louis completed an 8-8 season, summarising other aspects of season performance provides several arguments for ranking Carolina above St Louis. Carolina lost five matches by seven or fewer points and scored four or more touchdowns in five matches. St Louis, on the other hand, lost one match by seven or fewer points and scored four or more touchdowns in three matches. Also, Carolina scored 16 net points during the regular season and St Louis -73.

Bookmakers also appear to support our 2004 relative ranking of Carolina and St Louis. Specifically, Carolina was a seven point favorite when Carolina hosted St

Louis in week 14 (relatively late in the season) even though Carolina had an inferior win percentage. Indeed, throughout the 2004 season, bookmakers rated Carolina a stronger team than their win-loss record suggested. In seven of its 16 games, Carolina was the bookmakers' favorite even though either (a) Carolina's win percentage was inferior to its opponent's or (b) Carolina's win percentage was equal to its opponents and Carolina was the away team.

VI. Conclusions and Discussion

In contrast to the rigor and optimization involved in many SRSs, few league tables provide points for other than match outcome. The allocation of competition points used in most domestic and international rugby union competitions is a notable exception. In addition to rewarding wins and ties, this system awards bonus points for scoring more than a certain number of tries and losing by a narrow margin. Bonus points improve the accuracy of league tables and are easy for stakeholders to understand.

We designed an optimal bonus point system for the NFL by extending the work of Winchester (2008). As NFL schedules are unbalanced, our rating system employed (opponent) quality-adjusted wins, touchdown bonuses and narrow-loss bonuses. Three methods were used to average accumulated league points involving wins, ties and bonuses: time-varying average, moving average and exponential average. The predictive success of each model was optimized as to the minimum squared error between predicted and actual score differences. Both a touchdown-based bonus and a narrow-loss bonus were significant for NFL competition. Our preferred system, which is guided by mathematical optimality and practical implications, calls for four points for a win, two for a tie, one bonus point for scoring four or more touchdowns and one bonus point for losing by seven or fewer points. If adopted, such a system would make otherwise meaningless plays at the end of some games important as to league bonus points, which has implications for advertising revenue. By adding definition to league ratings, bonus points also assist the identification of teams in playoff contention. In summary, we conclude that there is considerable scope for the inclusion of bonus points in the NFL.

Our preferred allocation of competition points is similar to that used for most rugby union competitions. However, Winchester (2008) finds that the try bonus used in rugby union is not significantly correlated with team strength, but he does find a significant relationship between strength and a net-try bonus. This disparity is probably due to differences in how players function across the two codes. In rugby union, the same set of players is responsible for both offence and defence, so a team with a comfortable lead may relax defensively. In contrast, NFL teams have separate offensive and defensive squads and each player is carefully evaluated by coaches, making dominant teams less susceptible to defensive lapses. Ultimately, a weak NFL team is less likely to gain a touchdown/try bonus than a weak rugby union team.

Including bonus points in NFL standings is likely to be controversial for at least two reasons. First, the number of wins no longer dominates all other criterion in deciding NFL standings when bonuses are included. However, factors other than match outcome are considered by SRSs in many other competitions, and bookmakers' odds, which reflect punters' subjective ratings, regularly favor a team with fewer wins over a team with more wins.

Second, our procedure for choosing optimal league points did not allow bonuses to influence observed behaviour. We acknowledge this shortcoming but conjecture that bonus-induced changes in player and/or coach behavior will be small.

We also note several challenges in accounting for behavioural changes stimulated by bonuses. Addressing this limitation not only involves estimating the relationship between bonuses and actions, but also each team's ability at different tasks. Changes in how athletes train and/or recruitment policies motivated by bonuses further complicate such an analysis.

References

Associated Press (2005) A Super Bowl share is not so super paying, *nbcsports.com*, February 4.

Berry, S.M. (2003) College football rankings: The BCS and CLT, *Chance*, **16**(2), 46-50.

Boulier, B.L. and Stekler, H.O. (2003) Predicting the outcomes of National Football League games, *International Journal of Forecasting*, **19**, 257-270.

Clarke, S.R. (1993) Computer forecasting of Australian Rules Football for a daily newspaper, *Journal of the Operational Research Society*, **44**(8), 753-759.

Clarke, S.R. (2005) Home advantage in the Australian Football League, *Journal of Sports Sciences*, **23**(3), 375-385.

El-Bashir, T. and Heath, T. (2006) NHL's strong comeback marred by poor TV ratings, *Washington Post*, June 5.

Glickman, M.E. and Stern, H.S. (1998) A state-space model for National Football League scores, *Journal of the American Statistical Association*, **93**, 25-35.

Goode, M. (1976) Teaching statistics concepts with sports, Talk presented to the Joint Meeting of American Statistical Association, Boston, MA, August. (See also BudGoodeSports.com)

Harville, D. (1980) Predictions for National Football League games via linear-model methodology, *Journal of the American Statistical Association*, **75**, 516-524.

Harville, D.A. and Smith, M.H. (1994) The home-court advantage: how large is it, and how does it vary from team to team? *The American Statistician*, **48**(1), 22-28.

La Monica, P.R (2007) Super prices for Super Bowl ads, *CNNMoney.com*, January 3.

Leake, R.J. (1976) A method or ranking teams with an application to 1974 football, *Management Science in Sports*, North Holland Press.

Romer, D. (2006) Do Firms maximize? Evidence from professional football, *Journal of Political Economy*, **114**(2), 340-365.

Stefani, R.T.(1977) Football and basketball predictions using least squares, *IEEE Transactions On Systems, Man and Cybernetics*, **7**(2), 117-121.

Stefani, R.T. (1980) Improved least squares football, basketball, and soccer predictions, *IEEE Transactions on Systems, Man, and Cybernetics*, **10**(2), 116-123.

Stefani, R.T. (1987) Applications of statistical methods to American football, *Journal of Applied Statistics*, **14**(1), 61-73.

Stefani, R.T. (1997) Survey of the major world sports rating systems, *Journal of Applied Statistics*, **24**(6), 636-646.

Stefani, R.T. (1998) Predicting Outcomes, Chapter 12, *Statistics in Sport*, Oxford University Press.

Stefani, R.T. (1999) A taxonomy of sports rating systems, *IEEE Trans. on System, Man and Cybernetics*, Part A, **29**(1), 116-120.

Szymanski, S. (2003) The Economic Design of Sporting Contests, *Journal of Economic Literature*, **41**(4), 1137-1187.

Weisman, L. (2006) Expect NFL salary cap to keep going through the roof, *USA Today*, July 7.

Winchester, N (2008) Shifting the 'Goal Posts': Optimizing the Allocation of Competition Points for Sporting Contests, *Journal of Quantitative Analysis in Sports*: **4**(4), Article 1. Available at http://www.bepress.com/jqas/vol4/iss4/1 **.**

Zuber, R.A., Gandar, J.M. and Bowers, B.D. (1985) Beating the spread: testing the efficiency of the gambling market for National Football League games. *Journal of Political Economy*, **93**(4), 800-806.

TABLE 1

POINTS AND TOUCHDOWNS PER GAME AND WIN PERCENTAGES, 2003-2007

	(a) Intra-division		(b) Inter-division		(c) All matches	
	Home	Away	Home	Away	Home	Away
Score	21.9	20.0	22.9	19.8	22.5	19.9
Touchdowns	2.5	2.3	2.6	2.2	2.5	2.2
Win (%)	54.7	45.3	59.7	40.3	57.8	42.2

NOTE: Calculations exclude matches played at neutral venues.

TABLE 2
NFL AVERAGE REGULAR SEASON RATINGS

Team/Div/Conf	2003	2004	2005	2006	2007	All years
Buffalo	99.8	100.6	100.4	99.6	100.0	100.1
Miami	103.6	98.8	99.5	99.5	95.6	99.4
New England	103.9	108.0	106.0	106.1	113.1	107.4
NY Jets	101.4	102.1	99.1	99.3	98.2	100.0
AFC-East	102.2	102.4	101.3	101.1	101.7	101.7
Baltimore	100.7	102.3	99.4	103.7	100.4	101.3
Cincinnati	96.0	98.5	103.5	102.9	100.8	100.3
Cleveland	98.3	96.7	96.7	96.7	98.2	97.3
Pittsburgh	99.3	103.1	105.7	105.4	105.0	103.7
AFC-North	98.5	100.2	101.3	102.2	101.1	100.7
Houston	94.3	97.4	95.6	94.2	98.8	96.1
Indianapolis	103.1	107.4	109.3	106.1	108.8	106.9
Jacksonville	97.2	100.2	101.9	103.6	104.3	101.5
Tennessee	104.4	100.7	95.8	95.9	100.8	99.5
AFC-South	99.7	101.4	100.7	99.9	103.2	101.0
Denver	103.9	102.1	104.7	104.7	98.7	102.8
Kansas City	105.6	101.5	102.8	103.5	98.0	102.3
Oakland	101.1	95.5	97.1	93.8	93.9	96.3
San Diego	96.2	101.6	105.1	106.8	105.2	103.0
AFC-West	101.7	100.2	102.4	102.2	99.0	101.1
AFC	100.5	101.0	101.4	101.4	101.2	101.1
Dallas	98.2	96.7	99.5	101.5	104.3	100.0
NY Giants	99.4	96.9	100.2	101.2	100.4	99.6
Philadelphia	104.8	106.0	100.5	99.4	100.8	102.3
Washington	98.7	96.9	99.3	99.2	98.6	98.5
NFC-East	100.3	99.1	99.9	100.3	101.0	100.1
Chicago	96.0	97.1	98.8	104.7	99.9	99.3
Detroit	93.4	96.3	96.1	93.9	96.5	95.3
Green Bay	101.6	103.2	98.9	95.9	103.0	100.5
Minnesota	100.6	100.3	98.4	99.8	98.3	99.5
NFC-North	97.9	99.2	98.1	98.6	99.4	98.6
Atlanta	98.5	100.2	101.6	98.7	94.8	98.8
Carolina	99.8	101.4	103.0	101.4	98.3	100.8
New Orleans	99.5	98.4	97.5	99.6	98.8	98.8
Tampa Bay	106.9	99.8	100.5	96.7	99.1	100.6
NFC-South	101.2	100.0	100.7	99.1	97.8	99.7
Arizona	90.8	94.2	94.8	94.9	98.0	94.6
St Louis	101.8	99.2	96.1	96.5	94.6	97.7
San Francisco	100.4	95.4	91.1	93.0	94.2	94.8
Seattle	100.9	101.2	101.1	101.9	100.5	101.1
NFC-West	98.5	97.5	95.8	96.6	96.8	97.0
NFC	99.5	99.0	98.6	98.6	98.8	98.9

SOURCE: Authors' calculations based on exponential smoothing techniques set out by Clarke (1993).

TABLE 3
MODELING RESULTS

	TV		MA		EXP	
	(a)	(b)	(a)	(b)	(a)	(b)
Bonuses included?	No	Yes	No	Yes	No	Yes
Games in strength calculations	12	12	14	13	-	-
Touchdown partition	-	4	-	4	-	3
Narrow-loss partition	-	8	-	8	-	8
Home advantage (β^H)	1.98***	2.00***	1.93***	1.91***	1.99***	2.08***
	(0.63)	(0.63)	(0.63)	(0.63)	(0.65)	(0.65)
Inter-division home advantage (β^D)	1.38*	1.34*	1.46*	1.44*	1.42*	1.32*
	(0.80)	(0.79)	(0.80)	(0.80)	(0.80)	(0.79)
Net strength (β^S)	19.04***	19.51***	18.69***	17.18***	22.88***	20.29***
	(1.39)	(2.16)	(1.39)	(2.11)	(1.85)	(3.21)
Touchdown bonus (θ^{TD})		0.39***		0.53***		0.35**
		(0.15)		(0.18)		(0.17)
Narrow-loss bonus (θ^{LOSS})		0.49***		0.45***		0.46***
		(0.13)		(0.14)		(0.16)
R^2	0.16	0.17	0.15	0.16	0.16	0.17
Correct predictions	801	814	808	812	811	812
Percent correct	62.58	63.59	63.13	63.44	63.36	63.44

NOTE: ***, **, and * denote significance at the 1%, 5% and 10% significance level respectively. Robust standard errors for the TV and MA specifications are reported in parentheses. Standard errors for the EXP specification are derived using a Monte Carlo simulation. Sample size = 1,280.

TABLE 4

NFL PLAYOFF CLASSIFICATIONS CHANGES WHEN BONUSES ARE INCLUDED, 2003-2007

	2003	2004	2005	2006	2007
Baltimore				$z^* \rightarrow z$ 13-2-2 (56)	
Cincinnati			$z \rightarrow y$ 11-6-1 (51)		
Cleveland					$n \rightarrow y$ 10-4-4 (48)
Pittsburgh			$y \rightarrow z$ 11-5-4 (53)		
Indianapolis				$z \rightarrow z^*$ 12-5-3 (56)	
Tennessee					$y \rightarrow n$ 10-2-3 (45)
Dallas	$y \rightarrow n$ 10-1-1 (42)				
NY Giants			$z \rightarrow y$ 11-2-3 (49)	$y \rightarrow n$ 8-4-3 (39)	
Washington			$y \rightarrow z^*$ 10-5-5 (50)		
Chicago			$z^* \rightarrow z$ 11-2-1 (47)		
Minnesota	$n \rightarrow y$ 9-6-3 (45)				
Carolina		$n \rightarrow y$ 7-5-5 (38)	$y \rightarrow z$ 11-4-3 (51)		
Tampa Bay			$z \rightarrow y$ 11-1-3 (48)		
St Louis		$y \rightarrow n$ 8-3-1 (36)		$n \rightarrow y$ 8-5-3 (40)	

NOTE: *z* denotes division champion, * denotes direct qualification to the divisional round, *y* denotes wildcard qualifier, and *n* denotes non-qualifier. The first letter in each cell indicates the classification using a conventional system and the second shows the classification when bonuses are included. Numbers below classification change indicators convey wins, touchdown bonuses, narrow-loss bonuses and competition points (in parentheses) accumulated. Playoff positions for excluded teams are unaffected by the inclusion of bonuses.