

Abstract

The accuracy of NFL game predictions, as represented by the difference between the point spread and the actual outcome, was evaluated to determine if that accuracy improved over the course of a season. Data from the 1978 through 2013 seasons was analyzed and no significant change was found in the accuracy of the predictions over the course of the season. Additionally, data from the 2001 through 2013 seasons was analyzed, including a measure of how each team's season-long results differed from pre-season expectations. In games involving teams whose season-long results differed significantly from pre-season expectations, there was some evidence that early-season predictions of individual game results were worse than later-season predictions. Information about these teams' true qualities was quickly assimilated into individual game predictions; after the second week of the season predictions involving those teams were equivalent to predictions involving other teams.

Introduction

Convention wisdom would suggest that the accuracy of predictions for NFL games should improve as the season progresses. This would be due largely to additional information gained from the actual games played. This information could be in the form of both statistics and "softer" information gained from watching the games. One might expect that results of games at the very beginning of the season would be particularly difficult to predict. The performance of a team during the previous season might be a reasonable starting point for predictions regarding these games, but all teams undergo some turnover involving players and frequently coaches. Additionally, one may speculate about "intangibles" that could affect a team, but would only manifest themselves as the season goes on, e.g. a team's "chemistry".

A reasonable hypothesis would be that the accuracy of predictions would improve during the first few weeks of the season: the "live" games (as opposed to say, observations of practices and preseason games) provide significant additional information. The amount of improvement in predictions would then decrease, as each additional game provides less additional information. At some point in the season, the quality of the teams would be apparent, and the accuracy of the predictions would plateau. I investigated this by assessing the change in accuracy of gambling lines over the course of the NFL season.

The magnitude of such an effect could depend on the accuracy of the early (e.g. during preseason) assessment of a team's quality. That is, if the available information leads to an accurate early assessment of a team, perhaps no new information would be gained by watching that team play real games. To test this I used preseason gambling lines for teams' total wins, compared to their actual total wins. If this hypothesis were correct, the prediction errors in games involving teams whose season performance was similar to their predicted performance would tend to not improve during the season. Predictions of games involving teams that were better or worse than expected would show an improvement.

Note that this does not provide a mechanism for identifying such cases beforehand.

Literature Review

Ref 1 focuses on line changes (i.e. from when betting opens to when the games begins) but as an aside they run tests using the week and year as variables to predict the forecasting error (under the assumption that more information should be available in later years and in later weeks) and they did not find any significant results.

Methods

Data for the 1978 through 2013 NFL seasons were downloaded from <http://www.repole.com/sun4cast/data.html#dataprior> on September 4, 2017. Data associated with each game played included the date of the game, the teams involved, the spread, and the actual score.

Additional information for the 2001 through 2013 seasons was downloaded from [http://www.footballlocks.com/nfl_season_win_total_futures_lines_\[year\].shtml](http://www.footballlocks.com/nfl_season_win_total_futures_lines_[year].shtml) on September 4, 2017. For each team and each season, this data included the predicted number of wins at the beginning of the season (i.e. the over/under line for betting purposes) as well as the actual number of wins.

For each game, the difference between the predicted game result, as defined by the spread (i.e. the winning team and the margin of victory), and the actual result was calculated. I refer to this as the “spread error”. For instance, if team A was favored by 3 points and won by 6 points, the spread error for that game was 3 points. If team B won by 7 points the spread error for that game was 10 points. This error was our measure of the quality of each prediction.

For each game, the total number of games played by each team in that season, prior to the given game, during that season, was calculated. I refer to this as the “total games played” for that game. For instance, if the given game was team A’s seventh game of the season and team B’s sixth game of the season, the total games played for that game would be 13. Note that the total games played was generally close to twice the “week number” - 1 of the NFL season. However, due to regularly scheduled byes and special events (like the 2017 rescheduling of the Buccaneers – Dolphins game due to hurricane Irma) those numbers will frequently be slightly different. Total games played served as a proxy for the amount of information available from the occurrence of regular-season games.

For each season, for each team, the team’s actual number of wins was subtracted from the team’s predicted number of wins. This was used as a measure of the difference between the expected and actual quality of that team. For instance, if team A’s predicted number of wins was 8.5 and its actual number of wins was 6, the associated value would be 2.5, indicating the team’s results were worse than expected. I refer to this as the team’s “total wins error” for that season. For each game, the magnitude of the difference between the two team’s total wins error for that season

was calculated. I refer to this value as that game's "total wins error". For instance, if team A won 2 more games than predicted and team B won 2 more games than predicted, a game between them would have a total wins error of 0. If team C won 2 fewer games than predicted, a game between team A and team C would have a total wins error of 4. Total wins error was used to measure the "misinformation" that might be present at the beginning of the season, and could be corrected by additional knowledge gained during the season.

Data from the 1978 through 2013 seasons was used to evaluate the relationship between spread error and total games played by fitting a least-squares regression line. Additionally, results from approximately the first two games (total games played less than or equal to five) were compared to results from the rest of the season

The predicted total season wins by team was only available for the 2001 through 2013 seasons. Only those seasons were evaluated to evaluate the effect of total wins error on the spread error. Games were split into two groups, based on the total wins error of each game: a group of games in which the total wins error was less than or equal to four and a group of games in which the total wins error was greater than or equal to six. For each of these groups least squares regressions (as a function of games played) were calculated.

Those groups were each divided into two groups representing the beginning and the rest of the season. One group consisted of games with four or fewer total games played (essentially all games played in weeks one and two). The second group consisted of games with greater than 4 total games played. For these four groups the spread errors were compared to determine if there were statistically significant difference between their means.

Results

Figure 1 shows the absolute value of the spread error as a function of total games played for the 1978 through 2013 seasons. Note that a number of samples exceed the 30-point limit selected for the y-axis and are not shown. The linear least-squares regression line was: $\text{error} = 0.011 * \text{games_played} + 14.71$. The p-value associated with the hypothesis that the actual slope is zero was 0.38.

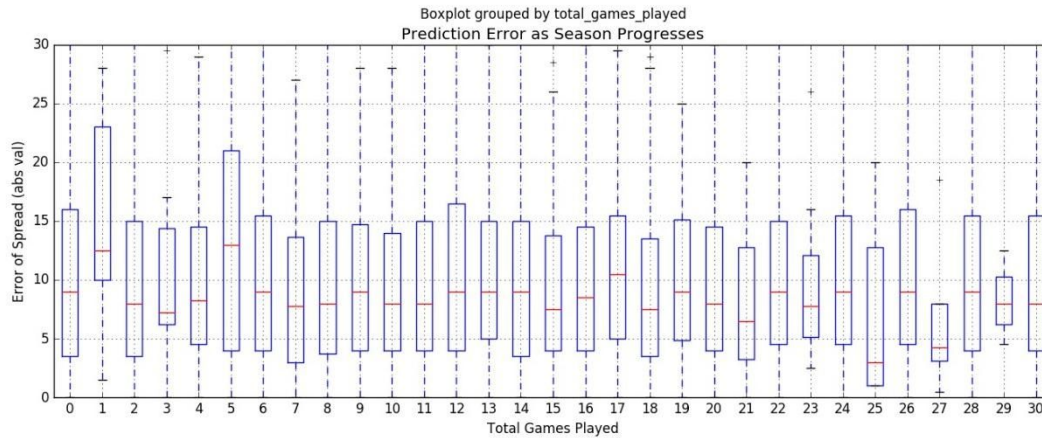


Figure 1 Over-under error by weeks

This data was divided into two sets, corresponding to games in which the total games played was less than or equal to four or greater than four. The mean spread errors for those groups are shown in Table 1. The P-value of a Ttest comparing the means was 0.75.

Table 1 Mean Line Errors, Early vs Late in Season

	Count; mean
Total games played ≤ 4	1587; 10.40
Total games played > 4	6,843; 10.47

As discussed in the methods section, total wins error is a measure associated with each game played. It measure the extent to which the pre-season expectations about the teams involved match their final results. Higher values indicate greater differences, e.g. one team performing much better than expected and the other team performing much worse than expected. Note that lower values will also tend to occur if both team's performance vary in the same direction, e.g. if both teams lost four more games than predicted, a game between them would have a total wins error of 0. A histogram of these values for the 2001 through 2013 season, is shown in Figure 2.

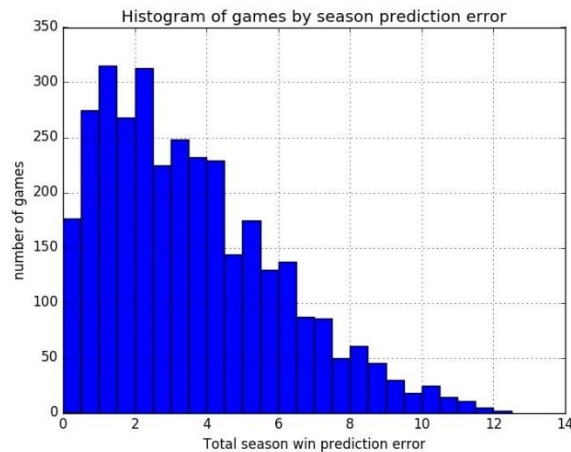


Figure 2 Number of games by season prediction error

I divided the games into two groups:

Games involving teams whose combined seasons were “expected”: 2,281 games in which the total wins error was less than or equal to four, and games involving teams whose combined seasons were “unexpected”: 574 games in which the total wins error was greater than or equal to six.

The mean line error of the “expected” group was 10.56; the mean line error of the “unexpected” group was 10.53. The P-value of the associated t-test was 0.95

For the “expected” group, the least-squares regression line had an equation of $\text{error} = 0.034 * \text{games_played} + 10.05$, $p\text{value} = 0.07$.

For the “unexpected” group, the least-squares regression line had an equation of $\text{error} = -0.096 * \text{games_played} + 11.96$, $p\text{value} = 0.01$.

Boxplots illustrating the variability of the line errors (actual line errors, not absolute values as are used elsewhere) for these groups are shown in Figure 1. For readability (and to eliminate entries with very few samples) the x-axis represents, roughly, the week of the season; it was calculated as $\text{floor}(\text{total_games_played} / 2) + 1$.

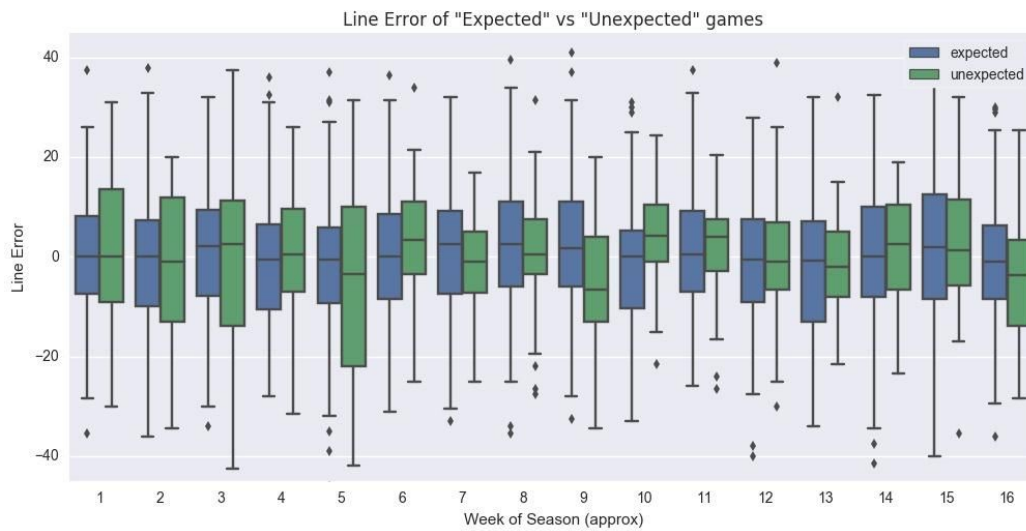


Figure 3 Boxplot of line errors by approximate week of season, expected vs unexpected games

Those groups of games were then categorized by the total number of games played, as shown in Table 2

Table 2 Count of samples in groups

Total number of games			
	Total games ≤4	4< Total games <28	Total
Expected	429	1852	2281
Unexpected	112	462	574
Total	541	2314	2855

The mean and standard deviation of the line error for these groups is shown in Table 3. Also shown are the results of four T-tests comparing combinations of the groups pair-wise.

Table 3 Line error by team groups and games played in season

Absolute Value line error: Mean (Standard Deviation)			
	Total games ≤4	4< Total games	Delta, Pvalue
Expected	10.00 (7.71)	10.68 (8.48)	0.68, 0.13
Unexpected	12.58 (9.06)	10.04 (8.22)	-2.54, <0.01

Delta, Pvalue	2.58, <0.01	-0.64, 0.14	
----------------------	-------------	-------------	--

Discussion

The initial results, analyzing all the 1978 through 2013 season, did not show evidence of any improvement in predictions as the seasons progressed. The positive slope of the linear model (somewhat inappropriate given the posited relationship between regular season games and available information for predicting outcomes) does not support the improvement in predictions as regular season games are played. The comparison between the first two games and the remaining season similarly did not support the idea that there is any special uncertainty during the beginning of the season.

The analysis incorporating pre-season expectations about each team suggested that there may be a subset of teams for which information is gained, and predictions do improve during the season. The two groups exhibited no significant difference in the line errors over the course of the entire season. However, the least squares regression line for the group with greater differences showed a statistically significant ($p=0.01$) negative slope (-0.096 per game), which would result in an 3 point improvement in predictions by the end of the season. The regression line associated with the games between teams performing closer to expectations was less statistically significant ($p=0.07$) and had a considerably smaller slope (0.034 per game).

Comparing the first two games of the season to the rest of the season demonstrated that predictions of early games involving “unexpected” matchups were significantly worse than early games involving “expected” matchups ($p<0.01$), as well as later games involving unexpected matchups ($p<0.01$). There was no statistically significant difference between later games involving expected matchups vs unexpected matchups.

This supports the hypothesis that there are teams (or more accurately, types of games) for which predictions improve during the season. Incorrect preseason assumptions about the strength of certain teams adversely affect predictions involving those teams. Results of the first two weeks’ games provide additional information that is incorporated into subsequent predictions for games involving those teams. For the remainder of the season there is then no evidence of differences in predictions involving those teams.

References

1. Xu, Steven. Forecasting Accuracy and Line Changes in the NFL and College Football Betting Markets. April 2013