Using Past Performance to Predict NFL Outcomes: A Chartist Approach

March 1997
This Revision: April 1997

David N. DeJong

Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260
dejong+@pitt.edu
(412) 648-2242

Abstract

A simple approach to predicting outcomes of National Football League games is demonstrated in applications to the 1995-96 and 1996-97 seasons. The approach amounts to a chartist strategy: it involves estimating team-specific probit models for predicting success or failure versus point spreads, using as explanatory variables own and opponent performance versus the spread in the previous week. Various strategies which trigger bets as functions of predicted probabilities of success are found to be profitable. Intraweek movements in betting lines are also found to be useful explanatory variables. The findings reflect negatively on the efficient markets hypothesis.

I thank Erick Elder, Steve Husted, Jean-Francois Richard, and Rick Tannery for useful discussions on this subject, as well as for financial contributions (via the office pool). I also thank Chuck Whiteman for constructive comments, and Kristin Anderson and Dilek Aykut for helping compile the data used in this study. Financial support from the NSF under grant SBR 9422828 is gratefully acknowledged. The usual caveat applies. In memory of my father.

Using Past Performance to Predict NFL Outcomes:
A Chartist Approach

David N. DeJong

I. Introduction

In setting point spreads on sporting events, gambling houses attempt to equate the flow of bets on both sides of the spread. Because bettors must risk $11 to win $10 (i.e., they must pay a ten-percent vigorish on losing bets), this point-setting rule ensures a profitable outcome for the house regardless of the outcome of the contest (with the exception of ties). Initial spreads issued by the house can thus be interpreted as "forecasts of the forecasts of bettors". Early movements in the spreads typically reflect inaccuracies in houses' forecasts; subsequent movements typically reflect the arrival of new information (e.g., injury updates); and differences between closing lines and final scores reflect forecast errors on the part of bettors.

The efficient markets hypothesis, applied to the gambling market for National Football League games, holds that point spreads are the best unbiased forecasts of actual outcomes. Under this hypothesis, it should not be possible to use past performance against the spread to predict future success or failure: to the extent that it is relevant, information regarding past performance should be embodied in current point spreads. The hypothesis does not seem to hold in this setting: using what amounts to a simple chartist technique, I demonstrate profitable strategies for predicting outcomes of NFL games.

The technique I employ involves estimating team-specific probit models for predicting success or failure versus point spreads. The models use as explanatory variables own and opponent performance versus the spread in the previous week, and movements in betting lines observed during the course of the week. Various strategies which trigger bets as functions of predicted probabilities of success are found to be profitable when applied to data from the 1995-96 and 1996-97 seasons. Due to the ten-percent vigorish charged by the house on losing bets, a success rate of 52.38 percent is required to break even. The strategies I consider have success rates ranging from 58 to 63 percent; these rates significantly exceed the break-even rate, both in an economic and statistical sense.

Several previous studies have examined the efficiency of NFL point spreads issued by Las Vegas betting houses. This work generally reflects positively on the efficiency of these spreads. Pankoff (1968) regressed winning margins on point spreads and a constant and found no exploitable biases using data from the 1956 - 1965 seasons, and Stern (1991) showed that differences between winning margins and point spreads measured from 1981 - 1986 are approximately normally distributed with zero mean and variance of 14.[1] Vergin and Scriabin (1978) reported finding 23 strategies among 70 competitors that generated winning percentages significantly greater (statistically) than 50 percent over the 1969 - 1974 seasons. However, Tryfos et al. (1984) showed that only three of these strategies were profitable after taking vigorish into account: i.e., only three strategies had success rates significantly greater than the break-even rate.[2] Finally, Zuber et al. (1985) reported a 59-percent success rate using a strategy based on predictions generated by a regression equation which models point spreads as a function of "fundamentals" such as number of wins per team, yards rushed, etc. The model was estimated using data from the first eight weeks of the 1983 season, and was then used to predict point spreads over the remaining eight weeks of the regular season. Discrepancies between actual and predicted point spreads were used to trigger bets; using a discrepancy of 0.5 points or more as a trigger, 60 of 102 bets turned out to be winners. But while the authors noted that this success rate is significantly greater than 50 percent, it is not significantly different from the 52.38 percent break-even rate (the p value associated with this test is 0.17 -- see Section IV for details on this test).

In sum, the literature on NFL point-spread behavior has generally supported the efficient markets hypothesis; here, the hypothesis is cast in a less favorable light.

II. The data

The data I consider are NFL point spreads and winning margins for the 1995-96 and 1996-97 seasons. Point spreads are Las Vegas betting lines as reported by the Associated Press; they were gleaned from the *Pittsburgh Post-Gazette* (*P-G*). Opening spreads are defined as the first available publication of spreads for upcoming contests. Closing spreads are defined as those published on game day. Differences in closing and opening spreads are defined as line movements.

In the 1996-97 season, on which I concentrated initially, the *P-G* published opening lines on Monday.[3] A histogram of line movements observed over the course of this season is illustrated in Figure 1a. Line movements were observed for 61 percent of all games played; 52 percent of these movements were by a mere 0.5 points. Such movements, though subtle, seem important. For example, consider the

[1]The means and standard deviations observed for the 1995-96 and 1996-97 seasons are (-0.9, 12.5) and (-0.43, 13).
[2] The profitable strategies amount to the location of cross-country arbitrage opportunities; they involve finding discrepancies in point spreads offered by bookmakers around the country on underdogs of five points or more. Badarinathi and Kochman (1996) found only one of these strategies to be profitable over the 1984 - 1993 seasons: bet on an underdog of five points or more if a two-point discrepancy can be found in favor of the underdog. They report a 56-percent success rate using this strategy.
[3] Over the course of the season, 18 games were listed on Monday as NL (no line). This typically occurred for games in which there was sufficient initial uncertainty about the status of one or more key players that bookmakers

spread on the Super Bowl, which opened at 13.5 in favor of Green Bay the day after the championship games, moved to 14 the following day, and did not move from that point forward. Bookmakers were willing to risk the possibility of a tie versus the spread in making this adjustment, a risk they did not face given the opening spread. So the value of this minor adjustment must have outweighed the risk of a tie, which was in fact realized: Green Bay won by exactly 14 points, an outcome that, according the to Associated Press, resulted in a decrease in winnings for Nevada bookmakers of approximately $5 million from the previous year. (Source: *P-G*, February 1, 1997.)

In the 1995-96 season, the *P-G* did not publish point spreads on Monday, so "opening" spreads were not available from this source until Tuesday at the earliest.[4] This delay matters: the histogram of line movements observed for the 1995-96 season illustrated in Figure 1b clearly contrasts with that illustrated in 1a. In the 1995-96 data, line movements were observed for only 51 percent of the total games played, and a χ^2 test of the null hypothesis that the two histograms were generated by the same underlying distribution rejects the null at the six-percent significance level. Initially, I found this difference in data sets disappointing, because it prevents a clean comparison of the forecasting performance of my procedure across data sets. However, this difference does enable a rough breakdown of line movements into errors in houses' forecasts of the forecasts of bettors, and movements in response to new information. Assuming that the former error is embodied only in the 1996-97 data, it follows that the ten-percentage-point difference in line movements observed across data sets is attributable to houses' forecast errors. This attribution is of course only an approximation, but it does suggest that houses are quite adept in forecasting bettors' forecasts.

Distributions of differences between winning margins and closing spreads observed over the two seasons are illustrated in Figure 2. These distributions are quite similar. As noted above, the means and variances computed over the 1996-97 season are -0.43 and 13; corresponding figures for the 1995-96 season are -0.9 and 12.5. Moreover, a χ^2 test of the null hypothesis that the two histograms were generated from the same underlying distribution fails to reject the null at virtually any significance level. While these histograms do not indicate obvious profit opportunities, the next section presents a simple approach for their discovery.

---

were unwilling to issue a spread. I included data on these games in my sample, but their exclusion yields similar results.

[4] I considered these data only after completing my analysis of the 1996-97 data; I did this to check the robustness of my findings for the 1996-97 data.

III. Charting success

My goal in evaluating the efficiency of NFL betting lines was to determine whether a simple backward-looking model was capable of outperforming Las Vegas spreads. In my view, the harder I had to search for an effective model, and the more complicated was the resulting model, the weaker would be the evidence (if any) I uncovered against efficiency. My search was a short one: the first set of team-specific models I considered yielded strong evidence against efficiency. The following subsection provides background for my choice of models; the subsequent subsection provides technical details.

Background

As an avid participant in recreational office pools (for entertainment only, of course), I have long relied on past performance to guide my weekly selections.[5] This has yielded mixed results: every year, it seems that some teams treat me well, while others wipe me out. Looking back at the 1996-97 NFL season suggested an explanation for this: many teams experienced extended streaks over the course of the season, thus rewarding my tendency to bet for last week's winners, and against last week's losers; at the same time, many others whipsawed over extended periods (e.g., won-lost-won-lost...), thus punishing my tendency.[6] (Explaining why such patterns coexist is difficult. Extended streaks could reflect adaptive expectations on the part of bettors; and whipsawing may be a manifestation of overshooting driven by an aggregate tendency to favor last-week's winners; but the compatibility of these explanations seems tenuous.) It occurred to me that different backward-looking models seemed appropriate for different teams. It also occurred to me that a formal statistical model had the potential to outperform my eyeball approach.

Besides past performance, I have also paid attention to line movements in making my weekly selections. The pool I participate in revolves around opening spreads, so I have interpreted line movements as signals of bargains generated by the market. I have found these signals valuable in competing against opening spreads; in specifying my team-specific models, I decided to investigate whether the signals were useful in competing against closing spreads as well.

The models

Since the outcome of a bet against the house can be thought of as a dichotomous random variable (equaling 1 if the bet wins and 0 if it loses), logit or probit models seemed well suited for fitting and

---

[5] I blame genetics: my father was a dyed-in-the-wool chartist.

[6] Notable teams in the former category were Indianapolis (eight-game losing streak); Green Bay (seven-game winning streak, five-game losing streak); and Carolina and Pittsburgh (six-game winning streaks). Notable teams

forecasting in this application. Given Stern's (1991) results on normality, I chose probit specifications, but logit specifications yield similar results. For the reasons given above, I estimated separate models for each team; the models consisted of a constant and three explanatory variables: own and opponent differences in winning margins and closing spreads from the previous game, and intraweek line movements. (I included only one lagged difference to keep the models simple, and to maximize the length of the forecasting window afforded by their use.)

I employed a dynamic forecasting algorithm in using these models to generate predicted probabilities of winning. The first set of probabilities I generated were for week eleven of the regular season; the models used to generate these probabilities were estimated using data observed over the previous ten weeks.[7] I then reestimated each model by updating the explanatory variables to include week-eleven observations, and generated a second set of probabilities. I repeated this process for the remainder of the season, including the playoffs and Super Bowl.

Before describing the algorithms used to process the resulting set of probabilities, two notes are in order. The first concerns the choice of the initial ten-week estimation window. Other choices are certainly possible, and one faces a clear tradeoff in choosing this width: shortening the window yields forecasts for earlier weeks, at the cost of a loss of observations available for estimating the models used to generate the earlier forecasts. The Washington Redskins are responsible for my choice of a ten-week window: I had to wait ten weeks before I could estimate their model, because their first eight dependent observations consisted of seven wins (speaking of streaks) and one bye week. I could have started forecasting in week ten by ignoring Washington and focusing on the remaining teams (doing so would have resulted in a five and two record for my leading algorithm in week ten), but I decided to begin in week eleven so that all teams could be treated symmetrically. The second note concerns the use of the dynamic updating algorithm used for reestimating the forecasting models. Use of this algorithm turned out to provide little value added over the use of the original models (estimated over the first ten weeks) over the entire forecasting horizon: it generated only one additional win, a result which speaks well for the stability of the models.

Use of these team-specific models yielded two predicted probabilities of success for each game: one for each contestant. I considered two strategies for triggering bets as functions of these probabilities. The first I will refer to as conservative: bet on a team if the predicted probability generated by its model is greater than 0.5, and the probability generated by its opponent is less than 0.5. The second I will refer to

in the latter category were New England and San Francisco (nine-week whipsaw streaks); and Denver and Seattle (six-week whipsaw streaks).

as aggressive: bet on a team if the predicted probability generated by its model is greater than the probability generated by its opponent's model.

In order to assess the marginal value of the information embodied in intraweek line movements, I generated a second set of predicted probabilities by dropping line movements as explanatory variables in the probit models, and repeating the process described above. So I considered a total of four betting strategies: conservative and aggressive, with and without incorporating line movements. The performance of these strategies is discussed below.

IV. Beating the house

As mentioned above, I initially applied my betting strategies to the 1996-97 season, and then applied them to the 1995-96 season to examine the robustness of my original findings. Table 1, fashioned after Zuber et al.'s (1995) Table 2, illustrates the payoffs generated by the four strategies I considered by presenting the results of a gambling simulation conducted for each strategy over the 1996-97 season. (To save space, simulation results obtained for the 1995-96 season are not tabled, but are summarized in the text.) The simulations involve betting $11 each time a bet is triggered. For each strategy, weekly wins and losses are reported, along with weekly and cumulative amounts bet, net winnings, and net returns. (Bets placed on games that resulted in ties are treated as nonbets.)

Consider first results obtained by including line movements as explanatory variables. Over the 1996-97 season, the conservative strategy had a 65.4-percent success rate, triggering 52 bets which netted $142 in net winnings (a 24.8-percent net rate of return). The aggressive strategy had a 58.4-percent success rate, triggering 113 bets and netting $143 in winnings (an 11.5-percent net rate of return). Similar results were obtained over the 1995-96 season: the conservative strategy had a 60.1-percent success rate (31 wins, 20 losses) and generated a 16-percent net rate of return, while the aggressive strategy had a 58-percent success rate (62 wins, 45 losses) and generated a 10.6-percent net rate of return. Weekly net returns exhibited high volatility over the course of the season, but cumulative net returns settled down quite quickly (approaching their ultimate levels within four to five weeks).

Consider now the results obtained by excluding line movements as explanatory variables. The success rate of the conservative strategy fell by five percentage points in the 1996-97 season, and the net rate of return it generated fell to 15.4 percent. The success rate of the aggressive strategy fell by three percentage points, and its net rate of return fell to 5.5 percent. The exclusion of line movements in the

[7] Typically, this estimation window yielded eight dependent observations by week ten: one observation was lost due to the use of lagged results as explanatory variables, and another was often lost because most teams enjoyed a bye week during this period.

1995-96 data set resulted in decreases in success rates of five and four percentage points for the conservative and aggressive strategies, and decreases in net rates of return to 7.2 and 2.3 percent. So in both data sets, the information content of line movements seems valuable: exclusion of these movements as explanatory variables is costly.

Information concerning the statistical significance of these findings is provided in Table 2. Classical and Bayesian measures of significance are reported. The Classical measures involve tests of the null hypothesis that the success rates reported above are significantly greater than 50 and 52.38 percent (i.e., the pure-chance and the break-even rates). These tests are conducted using Z statistics (differences between realized wins and wins expected under the null, measured in standard-deviation units computed under the null); critical values of the Z statistics are obtained using the normal approximation to the binomial distribution. In the table, Z1 denotes the test statistic computed for the pure-chance rate, and Z2 denotes the statistic computed for the break-even rate; the statistics were used to assess the season-specific and overall performance of each betting strategy. The Bayesian measures of significance are posterior odds ratios in favor of the null hypothesis that the winning percentages generated by each betting strategy are 55 percent, versus alternative hypotheses of 50- and 52.38-percent. The odds ratios were generated using the binomial distribution, and were computed using even prior odds.

Two features of Table 2 are particularly noteworthy. First, the two-year performances of both the conservative and aggressive strategies, applied to the predicted probabilities generated by the models which take intraweek line movements into account, provide sufficient evidence to reject both the pure-chance and break-even hypotheses. The conservative strategy generated 65 wins in 103 bets over this period -- a 63.1-percent success rate -- which leads to a rejection of the break-even hypothesis at the 3-percent significance level. Moreover, the posterior odds against the break-even rate are 2.8 to 1 in this case. Identical odds are obtained for the aggressive strategy, which generated 128 wins in 220 bets over this period, a 58.2-percent success rate (the break-even null is rejected in this case at the nine-percent significance level). Second, the value of considering intraweek line movements is again in evidence in Table 2. Exclusion of these movements leads to reductions in success rates over the two-season period of 5 and 4 percentage points for the conservative and aggressive strategies. As a result, posterior odds against the break-even rate are approximately cut in half for each strategy, and the null hypotheses that the success rates of these strategies are equal to the break-even and pure-chance rates cannot be rejected at the ten-percent significance level.

V. If you're so smart...

Wait until next year. Publicizing the performance of this simple, easily adoptable procedure now affords a truly challenging future test of market efficiency: if the procedure becomes well known and continues to succeed, the efficiency of the NFL gambling market will be cast further in doubt. In addition, I will have a satisfying answer to the question: If you're so smart, why aren't you rich?

References

Badarinathi, R. and L. Kochman (1996), "Football Betting and the Efficient Market Hypothesis," *The American Economist* 40: 52-55.

Pankoff, L. (1968), "Market Efficiency and Football Betting," *Journal of Business* 41:203-214.

Stern, H. (1991), "On the Probability of Winning a Football Game," *The American Statistician* 45: 179-183.

Tryfos, P., S. Casey, S. Cook, G. Leger, and B. Pylypiak (1984), "The Profitability of Wagering on NFL Games," *Management Science* 30: 123-132.

Vergin, R.C. and M. Scriabin (1978), "Winning Strategies for Wagering on National Football League Games," *Management Science* 24: 809-818.

Zuber, R.A., J.M. Gandar, and B.D. Bowers (1985), "Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games," *Journal of Political Economy* 93: 800-806.